UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

COMPUTATIONAL MODELS FOR NETWORK SOCIETIES

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS, MENCIÓN COMPUTACIÓN

FRANCISCO ANTONIO PLANA PERILLÁN

PROFESOR GUÍA:
ANDRÉS ABELIUK KIMELMAN

PROFESOR CO-GUÍA:
JORGE PÉREZ ROJAS

MIEMBROS DE LA COMISIÓN:
CLAUDIO GUTIÉRREZ GALLARDO
PABLO BARCELÓ BAEZA
BRUNO GONÇALVES

Este trabajo ha sido parcialmente financiado por CONICYT-PCHA/Doctorado Nacional/2016-21161085, infraestructura de supercómputo del NLHPC (ECM-02), Fondecyt 1200967, FB210017 Basal ANID, Instituto Milenio Fundamentos de los Datos, Facultad de Ciencias Físicas y Matemáticas Universidad de Chile.

SANTIAGO DE CHILE
2023

Resumen
Modelos computacionales para sociedades en red

Esta tesis trata sobre el modelamiento computacional de algunas hipótesis relativas a mecanismos de dinámicas sociales. La palabra *computacional* hace referencia al empleo de técnicas de ciencia de datos, con el objetivo de estudiar modelos expresados en un marco de redes, lo cual justifica la palabra clave *sociedades en red*, para referirse a los diversos fenómenos sociales representables a través de redes. El primer trabajo es un método sencillo para aproximar índices de centralidad de redes, llamado *QuickCent*, que está inspirado en heurísticas propuestas inicialmente para modelar procesos humanos de decisión e inferencia. El índice de centralidad que estimamos es la centralidad *armónica*, la cual es una medida basada en las distancias de caminos más cortos, lo que la hace infactible de computar en redes grandes. *QuickCent* es comparado con conocidos algoritmos de aprendizaje de máquinas en datos sintéticos, así como en redes empíricas. Nuestros experimentos muestran que *QuickCent* es capaz de hacer estimaciones que son competitivas en precisión con los mejores métodos alternativos testeados, obteniendo estimaciones con baja varianza del error a un costo de tiempo intermedio con una implementación sencilla, incluso con un conjunto de entrenamiento pequeño. Posteriormente discutimos sobre cómo QuickCent explota el hecho que en algunas redes, medidas de densidad local, pueden ser un buen proxy del tamaño de la región de la red a la cual un nodo tiene acceso. El segundo trabajo es un modelo de optimización de red inspirado por una dinámica de compartir comida que puede recuperar algunos patrones empíricos de redes sociales. Nos enfocamos en una formulación original de dos de los principales motivos discutidos en la literatura: la reducción del riesgo de hambruna individual, y el bienestar grupal o acceso igualitario al alimento, y mostramos que redes optimizando ambos criterios pueden exhibir una estructura de comunidades cohesionadas alrededor de cazadores, aquellos nodos que generan alimento. Adicionalmente, redes de bienestar óptimo se parecen a redes que promueven distribuciones de ingreso más igualitarias en juegos de regalos, y se obtienen distribuciones de reciprocidad que pueden ser consistentes con cómo el compartir se distribuye primero entre cazadores, y luego cazadores con sus familias. Estos resultados del modelo son consistentes con la visión que redes adaptadas para un uso óptimo de recursos, pueden haber creado el medio en el cual comportamientos prosociales evolucionaron, y en el que pueden ser potencialmente inducidos. Finalmente, los resultados sugieren que enfoques evolucionarios pueden beneficiarse de una perspectiva de distribución de recursos, la modelación de las necesidades de sobrevivencia, y la inclusión del nivel de análisis grupal.

Abstract

This thesis deals with the formal, computational modeling of some theories and hypotheses regarding mechanisms of social dynamics. The term *computational* refers to the usage of data science techniques, with the goal of studying models expressed in the framework of networks, which justifies the keyword of *network societies* to refer to the diverse social phenomena that may be represented through networks or graphs.

The first work is a simple and quick method to approximate network centrality indexes, called *QuickCent*, that is inspired in heuristics initially proposed to model some human decision and inference processes. The centrality index that we estimate is the *harmonic* centrality, which is a measure based on shortest-path distances, so infeasible to compute on large networks. We compare *QuickCent* with known machine learning algorithms on synthetic data and some empirical networks. Our experiments show that *QuickCent* is able to make estimates that are competitive in accuracy with the best alternative methods tested, achieving low error variance at an intermediate time cost with a simple implementation, even with a small training set. We discuss how QuickCent exploits the fact that in some networks, local density measures, may be a proxy of the size of the network region to which a node has access. These results show that simple heuristics are a promising line of research in the context of network measure estimations.

The second presented work, is a simple network optimization model inspired by a food-sharing dynamic that can recover some empirical patterns found in social networks. We focus on an original formulation of two of the main food-sharing drivers discussed by the anthropological literature: the reduction of individual starvation risk, and the care for the group welfare or egalitarian access to food shares, and show that networks optimizing both criteria may exhibit a community structure of highly-cohesive groups around special agents that we call hunters, those who inject food into the system. We have additionally obtained that optimal welfare networks resemble networks promoting more egalitarian income distribution in lab gift games, and also distinct distributions of reciprocity among hunters and non-hunters, which may be consistent with some empirical reports on how sharing is distributed in waves, first among hunters, and then hunters with their families. These model results are consistent with the view that social networks functionally adaptive for optimal resource use, may have created the environment in which prosocial behaviors evolved, and may also be potentially elicited. On the other hand, the model results have implications regarding the usual approaches adopted by evolutionary models. We claim that evolutionary approaches may benefit from a wider repertoire of assumptions including a resource distribution perspective, the modeling of survival needs, and the explicit inclusion of the group level of analysis.

Dedicada a quienes son voluntarios frente a su vida.

Acknowledgments

Esta tesis y todos mis logros no habrían sido posibles sin el apoyo perenne de mi familia. Estoy profundamente agradecido por la entrega incondicional de mis padres, Carmen y Sergio, a lo largo de toda mi vida. Vuestro amor y dedicación es una energía que inspira todos mis esfuerzos. Mis hermanas Lucía, Antonia, y Javiera, son mis maestras personales de la generosidad, empatía y pasión, muchas gracias por vuestra sabiduría. Mis abuelos Checa, Toño, Marta, Aurora y Miguel siempre han manifestado, o manifestaron, una alta estima por el conocimiento y la educación en general, por lo que esta actitud fue seguramente un importante ingrediente explicando este documento y varias de mis decisiones. Mi querida familia extendida formada por mis tías Cecilia, Quena, mis tíos Miguel, Pepe, Poque, y mis amigues Sebastián, Cristian, Javiera, Josefa, Ginette, Paloma, Cristián, Leslie, Beatriz, Raimundo, Mónica y Cristian, y sus familias, siempre me han dado apoyo y mostrado gran respeto por mi trabajo, también aprecio eso.

Un soporte clave para este trabajo es debido a quien fue mi guía durante la mayor parte de la tesis, Jorge. Estoy en gran deuda por toda su paciencia, comprensión, enorme libertad que siempre me dio para emprender este extraño viaje de los estudios de doctorado. Él nunca dudó del valor de lo que yo podía entregar, lo cual es algo que tuve que aprender para poder terminar esta investigación, el valor de la autoconfianza. Yo admiro no solo su habilidad para investigar, sino que también su ética y amplitud de criterio para apoyar actividades que van más allá de los deberes usuales de la investigación institucional, esta tesis también se debe a este atributo de él. Mi supervisor actual, Andrés, pese a venir a guiar la última parte del trabajo, ha hecho de todas formas, gracias a su lectura crítica, importantes contribuciones en conceptualizar el modelo de compartir comida y la versión actualizada de QuickCent, así como en su proceso de escritura y publicación. Además de ser un investigador bien talentoso, él cumple sus compromisos, de este modo ha ayudado a llevar este trabajo por el camino de menor resistencia, así que también estoy agradecido de él.

Finalmente, esta tesis también se debe de forma importante al ambiente de cercanía, y al mismo tiempo, de trabajo riguroso desarrollado por las personas del Departamento de Ciencias de la Computación de la Universidad de Chile, incluyendo estudiantes, y todo el equipo funcionario académico y no académico, todes hacen un gran trabajo para que esto suceda. También es necesario agradecer la beca para estudios de doctorado nacional de ANID (ex-CONICYT) [CONICYT-PCHA/Doctorado Nacional/2016-21161085] por sustentarme durante estos años, a la beca de doctorado de la Escuela de Postgrado y Educación Continua de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile por apoyarme en la última etapa de estudios, y al Laboratorio Nacional de Computación de Alto Rendimiento (NLHPC), sin Leftraru las redes óptimas de compartir comida no habrían visto

la luz.

Powered@NLHPC: Esta investigación/tesis fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02).

Por último, no puedo dejar de mencionar el importante rol en este trabajo de quizá el experimento social más grande de la historia, la WWW. Este trabajo le debe mucho a todas aquellas personas que hacen su parte en sostener el conocimiento libre y compartir contenido valioso de forma gratuita. Esto puede bien ser, en retrospectiva, una de las mayores revoluciones en la historia de la humanidad. Un movimiento distribuido y auto-organizado de la humanidad para liberarse de sus prisiones auto-construidas.

Table of Content

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Topic and contents of this thesis

This thesis deals with the formal, computational modeling of some theories and hypotheses regarding mechanisms of social dynamics. The term *computational* is used in a wide sense, referring to the use of modeling strategies coming from either physics, mathematics, statistics, computer and social sciences like economics or sociology. The underlying meta-assumption is that the approach of computational modeling is a powerful methodology to advance the discussion of ideas that, otherwise, would be difficult to define more precisely or disentangle. A common feature of the models of the thesis is the usage of networks as formal framework, which justifies the keyword of *network societies* to refer to the diverse social phenomena that may be represented through networks or graphs. The field of networks science is comprised mainly of the studies of complex systems formalized as networks, which is the field known as complex networks.

The two research lines can be methodologically described as the use of data science techniques with the goal of studying optimization models inspired by social phenomena. The data science techniques used along the thesis comprise the use of computational statistics [67], clustering analysis [275, 10], machine learning [189, 48, 261] and meta-heuristics [69, 91]. The use of these techniques corresponds to another meaning of the term *computational models*. Now, while the food-sharing model is an explicit network optimization model, QuickCent is a heuristic that attempts to minimize the approximation error, but this definition is not given as a formal optimization problem, so we refer to it as an implicit minimization. Thus, this thesis is directed to the interdisciplinary community at the intersection of quantitative machine learning and network science.

Next, we give the motivation for the problems addressed in the thesis. In Chapter 3, a simple heuristic is proposed to give quick estimates of harmonic centrality [193]. The centrality measures are an attempt to quantify the intuitive notion of the importance of an actor given by its location in a social network, and these indexes have a long history of applications of graph theory in the social sciences [40, 285]. The harmonic centrality is a centrality index expensive to compute, and sensitive to either the size of the actor's

group and the density of connections of its local neighborhood, according to one axiomatic characterization of centrality [40]. Our heuristic is able, under certain conditions, to give accurate and quick estimates of harmonic centrality using the, cheap to compute, local in-degree centrality. To the best of our knowledge, our work is the first practical application based on this axiomatic characterization.

On the other hand, the model of Chapter 4 is an attempt to recover the structural network implications of human food-sharing, a phenomenon argued to be an important evolutionary force shaping several social behaviors and features of the foraging niche [124, 163, 245]. The patterns of human food sharing stand out by its unique complexity, which extends well beyond infancy lactation to the whole life and across adults and families [124]. These patterns may have emerged as a response to a greater offspring dependence associated with the development of a larger brain requiring a nutrient-rich diet, or more difficult feeding strategies, in contrast to other wild primates having more predictable diets [145, 124, 154]. Our model provides a novel formulation of several motives for food-sharing proposed in the evolutionary anthropology literature [151], and develops the hypothesis that the modular character of food-sharing networks, a characteristic feature of social networks [212], may be an optimal network feature when minimizing the objectives of individual and collective starvation risks.

The chapters of the thesis correspond to the following texts or contributions.

- Chapter 2 is a non-exhaustive survey of the literature of complex networks, with an orientation towards the study of network societies, or communities within complex networks. This survey is written with the goal of providing the necessary context for subsequent chapters, and is partially based on the survey document written as a requirement for the Qualifying exam of the PhD program. Connections to later sections of the document are: the estimation of power-law distribution parameters (Section 2.3.3), the general preferential attachment growth model (Section 2.3.4), as well as the insight that this mechanism is a better description for information networks (Section 2.3.6) mentioned in the QuickCent model (Chapter 3). The other connection to later chapters is the section on the distinct hypotheses for communities modeled as networks (Section 2.4), since the Food-sharing model (Chapter 4) may be seen as a new hypothesis of the origin of network societies, nearer to the risk-sharing theories (Section 2.4.3). Annex A is also part of this chapter, and is a formal argument from the literature given to highlight the importance of the behavior of the moments of the power-law distribution.

- Chapter 3 is based on the following conference paper.

  Plana, F., & Pérez, J. (2018, December). QuickCent: A Fast and Frugal Heuristic for Centrality Estimation on Networks. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 238-245). IEEE.

  The paper describes a proof-of-concept heuristics designed to deliver fast and accurate estimates of the harmonic centrality, which follows a power-law distribution on some networks. The chapter is an updated version of the paper, presenting important changes in relation to it. These changes comprise some implementation details of the heuristics (computation of proportions vector, the use of the median of the lowest centrality values,

in Section 3.3), the study of the general preferential attachment model (Sections B.2, B.3, 3.4.1), and the assumptions of QuickCent (Section 3.4.3), as well as the experiments on empirical networks (Section 3.4.4) and the discussion (Section 3.5). Annex B is also part of this chapter, which gives details on the practical estimation of the power-law parameters, and the experiments for checking the fulfillment of assumptions of QuickCent by the distinct networks shown. For the time of this writing, we have submitted this chapter to the journal *Computing*.

- Chapter 4 is based on the following journal paper.

  Plana, F., Pérez, J., & Abeliuk, A. (2022). Modularity of food-sharing networks minimises the risk for individual and group starvation in hunter-gatherer societies. To appear in *PloS one*.

  The paper proposes a network optimization model designed to contrast the hypotheses of minimization of individual versus group starvation risk in the context of hunter-gatherers' food-sharing, by exploring the optimal networks under each assumption, and comparing their relative potential of recovering empirical network structures such as the presence of modules or communities. The content of the chapter is practically the same as the accepted paper. This research has a first publication as the following conference poster available in the Annex C.

  Pérez, J., & Plana, F. (2020). Food sharing gave birth to social networks. In *CogSci*.

  Annex D includes implementation details of several components of the model such as the estimate of the probability of eating, the structure of the simulations, the evolutionary algorithms, clustering procedures and decision trees. Annex E displays a series of analysis made later to the paper submission, with the goal of giving an initial answer to the following question. Are the optimal networks from this food-sharing model able to recover the scale-free property? It is an important question considering that this property, together with the community structure, are the main themes of this thesis, and this model was designed with the goal of studying only a hypothesis for community structure in a specific type of social networks.

In the next sections, we give more context to each of these chapter contributions.

1.2 The complex network approach to network societies

Here we present the argument of Chapter 2, which provides a non-exhaustive survey on the topic of modeling of social networks within the frame of complex networks, with tools and concepts that are used in the later contributions of this thesis. In addition, the chapter also discusses about the specific characteristics of social networks to pose a critical claim on the motivations of some developments in the science of complex networks. Namely, we present converging evidence [53, 148, 21, 243, 143, 277] suggesting that pure social networks would be at most weakly scale-free (Section 2.3.6), which introduces a nuance to those modeling efforts seeking to simultaneously reproduce several patterns of complex networks (Section 2.3.5), in the sense that these patterns would have different importance and interplay depending on

the specific type of network studied. This statement is a unifying theme of this thesis, since the heuristic presented in Chapter 3 is particularly suited for networks generated by preferential attachment, which would be a better descriptive mechanism for information or knowledge networks [210], such as the web, citation or collaboration networks, rather than pure social networks like friendships. For this reason, if the goal is to study social networks or network societies, it is desirable to bear in mind the explicit social mechanism under study. Section 2.4 reviews several mechanisms of social dynamics engendering communities, and the contribution of Chapter 4 can be considered another work in this line. We next review in more detail the arguments displayed by the survey.

The survey starts in Section 2.2 by defining the central features characterizing complex networks, the *scale-free* property, or the power-law distribution of the degree or the number of connections by node, and the *small-world* property, which corresponds to the short distances connecting nodes in a network, as well as the features defining *network societies*, namely, the positive degree correlations or, equivalently, the presence of *communities*, or densely connected groups. Then, Section 2.3 makes a brief review of the contemporary field of networks science, starting with some of the first and most studied models of preferential attachment (PA) and other equivalent mechanisms (Section 2.3.1), to show next the shortcomings of these models to simultaneously recover all the features of complex networks, as well as other criticisms (Sections 2.3.2, 2.3.3). It follows with a bridge section (Section 2.3.4) of the general PA growth, which motivated the study of important networks in the field like citation, collaboration, and Internet. The survey then passes on later works (Section 2.3.5) that, looking for unifying the complex networks traits that the first models are not able to reproduce, propose mechanisms inspired on citation networks and the hierarchical network structure. Section 2.3.6 introduces a nuance to these unifying aspirations, at least for the study of social networks, in the sense that it reviews converging evidence from distinct works pointing to the idea that social networks would have only a weak scale-free character. Finally, the chapter is ended by Section 2.4 where distinct kinds of models of communities in networks are reviewed.

1.3 QuickCent model

In Chapter 3 it is reviewed the QuickCent model, which is a heuristics to approximate the harmonic centrality from the knowledge of the in-degree, and in general, to approximate any magnitude distributed according to a power-law, related via a monotonic function to another variable from which the estimate is computed. The merits of the heuristics is that, despite being very simple, its estimates are competitive in accuracy with some of the best alternative machine learning methods tested, achieving low error variance and intermediate time costs, with a naive implementation, even when trained with small training sets of 10% of the total vertex set, of either synthetic or empirical networks of around 10000 nodes (Sections B.3, 3.4.1, 3.4.2 and 3.4.4). There are antecedents in the literature of distinct simple heuristics having a performance comparable or superior to more complex models, such as medical decision making [13], predicting the outcomes of sport matches [253], geographic profiling [259], or forecasting future purchase activity of customers [293]. There is also a contest [292] of machine learning algorithms and simple estimation heuristics on 99 real

world datasets from distinct sources, finding that a variant of QuickEst, the antecedent of QuickCent (see the next paragraph), is one of the best performing methods. Our proposal differs from previous works in that it requires only one predictor from which the estimates are built, and it is a general method working when its stated assumptions hold.

QuickCent is a generalization of QuickEst [132], a heuristic proposed to represent the processes underlying human quantitative estimation. QuickEst is a very simple heuristic that relies on sequences of binary clues associated, in our context of centrality approximation, to nodes in a network, where the value of a clue is an indicator of the presence or absence of an attribute signal of greater centrality for a node. The method simply finds the first clue with negative value (absence), and it outputs the mean of points with a negative value on this clue. For this reason, QuickEst is part of the cognitive model of human inferences and decision making named *Fast-and-frugal* heuristics [109], in the sense that they do not search for all the available information, in contrast to traditional structural linear models employed in social judgement theory [82], which are compensatory, that is, they integrate all the information available by weighting the distinct components. On the other hand, QuickEst has been argued to present a *negative bias* [132], in the sense that it is a negative clue (or absent attribute) what stops this heuristic. Thus, a distribution such as the power-law where most values are small (with mostly negative cues) and only few high values exist (with mostly positive cues), would provide an optimal context for the performance of QuickEst. This is an example of what has been called as *ecological validity* [132], which is the fit between the cognitive model and the environment in which it operates, which would be one of the factors explaining the success of these simple heuristics. The reasons explaining why and when these simple heuristics work is an active research problem [49, 134].

Finally, there is an observation to make regarding the study case we address of approximating an expensive index like the harmonic centrality, sensitive to either neighborhood density and size of the belonging group [40], by cheap local density measures such as the in-degree. As it is discussed in Section 3.5, the preferential attachment mechanism, which is the model used to generate the synthetic networks to test QuickCent (Sections B.3, 3.4.1, 3.4.2), may be produced by processes guided by the local network structure such as meeting friends of friends [143, 277], because higher degree nodes are more likely to be found since more paths lead to them. That is, the local density could indeed reflect the access to bigger parts of the network. In summary, it would be ecologically valid to estimate group size from local density in preferential attachment networks (Section 2.3.4), which would explain the high accuracy achieved by QuickCent on the tested synthetic datasets.

1.4 Food-sharing model

In Chapter 4 it is reviewed the Food-sharing model, where the optimal networks resulting from distinct objective functions are contrasted, with the goal of determining their power to recover social networks patterns such as the presence of communities or reciprocity. The model is formulated as a global optimization, very much in line of the proposal by Ferrer i Cancho and Solé (2003) [57], discussed in Section 2.3.1. However, the joint optimization of two goals has been formulated in our model with the concept of Pareto optimality, see Section

4.2.4, rather than taking the common approach of the convex combination of goals, since it has been studied that this formulation may fail in obtaining the Pareto solutions located in the concave region of the Pareto front [99]. Now, while it has been argued that either the scale-free property or the clustering of empirical networks is more likely the product of local optimization processes [277, 263, 74], see Section 2.3.5, our food-sharing model makes no claim about the dynamics or trajectory leading to these optimal networks. It simply presents what type of food pathways structures are likely arising under the assumption that these are optimally organized according to one functional goal or another. With this, the model does not look for neglecting the series of complex phenomena converging in food-sharing, as it is discussed on Section 4.1. These exchanges ensuring the most elemental subsistence, have been suggested to provide the material basis for inter-generational accumulation of cultural innovations and collective memory [201].

The contributions of the Food-sharing model are the following. First, the ability to recover some empirical social networks patterns. Networks optimizing both criteria may exhibit a community structure of cohesive groups under stringent conditions of food supply, see Section 4.3.3. These networks are resemblant to the food-sharing networks observed by Dyble et al (2016) [89], where there are cohesive groups corresponding to households provisioned by an adult couple, and a set of households forms a cluster with a small number of inter-household connections. We have additionally obtained that optimal welfare networks, where each hunter is connected to one big, homogeneous and dense group of non-hunters, see Sections 4.3.1, 4.3.2, resemble social networks that promote more egalitarian income distribution in a lab gift game reported by Chiang (2015) [64]. This model result is consistent with the view that social networks structured originally according to ecological considerations, may have created the environment in which prosocial tendencies and equity response elicitation evolved [144, 139]. On the other hand, distinct distributions of reciprocity are obtained for each optimization regime, see Section 4.3.3, which may be consistent with the empirical finding [153] that often the sharing occurs in waves: first among hunters, and then hunters with their families. This model result gives a broader picture of the usual notion that reciprocity is driven by the minimization of food production uncertainty [155].

In second place, the model results have implications regarding the usual approaches adopted by evolutionary models. Typical evolutionary models of food-sharing [151] rely on the assumption of maximization of a function of subjective preferences, usual in economic models of network formation [156], which highlights an exclusive role of individual choice and may impose high cognitive requirements, see Section 4.4. The goals that our model looks for minimizing may in fact be interpreted as a reformulation of the usual motives given as the causes of food-sharing, where the individual starvation risk would map to the reciprocal altruism, or tolerated theft, while the group starvation risk would correspond to the need motive, see Section 4.1. The problem with individualistic evolutionary accounts is that reciprocal altruism is evolutionarily adaptive, while need is not [257], see Section 4.1. The results of our model show that the need motive is an equally important and powerful principle to theoretically explain network structures observed in empirical food-sharing networks, such as communities. We claim that evolutionary approaches may benefit from a wider repertoire of assumptions including a resource distribution perspective, the modeling of survival needs, and the explicit inclusion of the group level of analysis, see Section 4.4. Finally, our model makes some methodological contributions. Our model is based on an original formulation

of starvation risk relying upon the expected number of success runs of a certain length in a sequence of trials to implement the hunger need over time, see Section 4.2.3. The clustering stage of the pipeline of analysis of the optimal networks dataset, comprised of the application of tSNE algorithm [275] followed by OPTICS [10], see Section 4.2.5, implemented in order to to be flexible enough to accommodate clusters of arbitrary shapes and distribution of points, may also be of interest.

Chapter 2

The complex networks approach to network societies

Abstract

This chapter aims to give a consistent survey of those topics in the interdisciplinary literature of networks science, that will serve as elements of context for the rest of the thesis work. A central idea of the survey is that while complex networks posses, in general terms, several common patterns, network societies or networks having communities display their own specificity of importance and interplay of these patterns. This idea is also a unifying theme of the whole thesis document, since Chapter 3 discusses that the heuristics introduced therein may be better suited for information networks such as citation networks or the World-Wide Web, rather than pure social networks, and Chapter 4 is an explicit mechanism of a type of social dynamics engendering communities. Thus, some key points of the survey are the emphasis on modeling and fitting of power-law networks, basis for Chapter 3, and models of communities, which are the defining feature of network societies, the representation of social dynamics in terms of complex networks, providing the foundation for Chapter 4. The focus of the chapter is oriented mainly to study different abstract mechanisms, simple hypothetical principles of network generation proposed to recover regularities in empirical networks, together with the study of social dynamics in terms of the complex networks formalism. The field of network science is immense, and this chapter does not pretend to be exhaustive at all. Topics covered include central features of complex networks, basic jargon about power-law distributions, a brief review of the contemporary field of networks science comprising the first models, criticisms in their merit to recover and fit empirical features, and later works that address detected shortcomings in the first models. The chapter is ended by a section of models of communities in network societies.

2.1 Introduction

This chapter aims to give a consistent survey of those topics in the interdisciplinary literature of networks science, that will serve as elements of context for the rest of the thesis work. A central idea of this survey is that while complex networks posses, in general terms, several common patterns, network societies or networks having communities display their own specificity of importance and interplay of these patterns. This idea is also a unifying theme of the whole thesis document, since Chapter 3 discusses that the heuristics introduced therein may be better suited for networks incorporating an information component (Section 3.5), rather than pure social networks, and Chapter 4 is an explicit mechanism of a type of social dynamics engendering communities (Sections 4.3.3, 4.4). Thus, some key points of the survey are the emphasis on modeling and fitting of power-law networks (Sections 2.3.3, 2.3.4), basis for Chapter 3, and models of communities in networks (Section 2.4), which are the defining feature of network societies, providing the foundation for Chapter 4.

The focus of the chapter is oriented mainly to study different abstract mechanisms, simple hypothetical principles of network generation proposed to recover regularities in empirical networks, together with the study of social dynamics in terms of the complex networks formalism. The field of network science is immense, and this chapter does not pretend to be exhaustive at all. Therefore, omissions are made on important topics. Some of them are: tractable models looking for appropriate null models to infer typical behavior on some variable [65, 204, 194, 280, 55, 38], models looking explicitly for fitting or reconstructing real-world networks [168, 181, 218], or motifs [147]. The review is not even exhaustive on the generative mechanisms topic, since, for example, an important kind of models based on geometrical considerations or lattices [172, 299, 249] are left out. The field of network science is a mature one, and the interested reader may consult the books by Caldarelli (2007) [54], Barabási (2016) [18], Newman (2018) [207] or Jackson (2010) [141].

Thus, the rest of the chapter goes as follows. The chapter starts in Section 2.2 by defining the central features of complex networks, the traditional modeling approach of the random network, and basic jargon about power-law distributions. It is followed by Section 2.3 where a brief review of the contemporary field of networks science is presented. It reviews the first and most studied models, criticisms to their merit in recovering and fitting empirical features, and to other aspects. The review then passes on later works that, looking for unifying the complex networks traits that the first models are not able to simultaneously reproduce, propose mechanisms inspired on citation networks and hierarchical network structure. Next, a nuance is presented to the unifying program, since evidence showing that social networks would not display all of the complex networks patterns is reviewed. Finally, the chapter is ended by Section 2.4 where distinct kinds of models of communities in networks are reviewed.

2.2 Basic Concepts

In this section we give an introductory background with the building blocks of what follows in the chapter and the rest of the document. It starts with a brief introduction of the meaning and properties of complex networks and network societies. Then, it continues with a brief

review of the classical random network models, and how their limitations motivated the arising of the power-law network models. The section closes with some basic concepts and facts of the power-law probability distribution.

2.2.1 Features and relevance of complex networks

Complex networks are the representation, in terms of a network or graph of interacting *vertices* or *nodes*, of complex systems coming from nature and society. They are the subject of study of the interdisciplinary field of network science, which has experienced a large growth in the last two decades. Examples of these complex systems studied range from distinct maps from cell biology [4], the potential energy landscape of an atomic system [86], technological systems such as the internet [96] or social networks such as scientific collaborations [21]. In spite of the diversity of studied systems, most of these networks present common statistical features that suggest the existence of common mechanisms of network generation. These features are the *scale-free* or *power-law* topology [20], the fact that the *degree* or the number of connections each vertex has follows a power-law distribution, and the *small-world* property [5, 24, 286], which is the joint presence of high *clustering*, or the probability that vertices connected to some vertex are in turn connected, and small *diameter*, a measure of the shortest distance between any two vertices.

Another feature that may be measured is the *assortativity*, or the correlation between connected vertices on some property like the degree. Thus, it has been reported that social networks are assortative, that is, they have positive degree-degree correlations, and that biological or technological networks are disassortative, that is, they have negative degree-degree correlations [210]. These patterns have been explained by arguing that social networks have *communities*, or groups of vertices more densely connected among them than with the rest of the network, since this feature is able to account both for the high clustering and the assortativity [212]. On the other hand, a principle of maximum entropy has been invoked to show that, in the absence of additional knowledge about the generative mechanism, scale-free networks are likely disassortative [150]. It is this distinctive feature of having communities by which we will to refer to social networks as *network societies*.

These patterns found in complex networks confer them special properties when dynamical processes are run over these structures. Scale-free networks provide a better environment for the evolution of cooperation in public goods games than regular graphs [252], that is, graphs with constant degree. Another relevant property of scale-free networks is the absence of an epidemic threshold for the SIS model [221]. This threshold corresponds approximately to the number of infections each infected produces, and imposes a limit on the evolution of the epidemic, whether it turns out to be endemic, or it dies out. The absence of this threshold means that any virus, whatever infection rate it has, can spread over the network. Scale-free networks show also a high level of tolerance against random vertex removal, which is reflected in a smooth increase of the diameter, but display a rapid increase in the diameter in case of removal of the most connected nodes, sorted by their degree [7]. On the other hand, networks with community structure promote the evolution of collective fairness in multiplayer ultimatum games [251] and inhibit the spreading of opinions and hinder the effectiveness of committed agents in a binary naming game [202]. Additionally, on epidemic processes, they

may exhibit a global threshold which sets a critical value of the diffusion rate below which the epidemic is not able to spread to a macroscopic fraction of subpopulations [72].

In the next subsections, we will review the first modeling attempts of complex networks, and the basics of some of the most known current models.

2.2.2 Random network model

One of the first and main network models is the Erdös-Rényi graph [92]. It was proposed by Solomonoff and Rapoport (1951) [260], and then independently rediscovered and rigorously studied by Paul Erdös and Alfréd Rényi in a series of seminal works [92, 93, 94]. It is a very simple model that, however, has been greatly influential to the study of networks due to the intuitions it provided. In this model, we have a graph of n vertices, and each possible pair of vertices is connected with probability p. This defines a set of undirected graphs we name as $G_{n,p}$. The model by Erdös-Rényi is parametrized by the number of edges in the graph. Here, we follow the equivalent formulation by Gilbert (1959) [110], which is parametrized by the connection probability p. For these graphs, the expected number of edges is

$$\mathbb{E}(|E|) = \frac{pn(n-1)}{2},$$

and the average degree is

$$\langle \deg \rangle = \frac{\mathbb{E}(\sum_i^n \deg_i)}{n} = \frac{\mathbb{E}(2|E|)}{n} = p(n-1) \simeq pn.$$

One of the aspects of real networks that can be reproduced by Erdös-Rényi random graphs is the empirical observation that most vertices in a network are connected by short paths. In this model, the mean number of nodes at distance d is given by $\langle \deg \rangle^d$. Then, the maximum distance in the network d_{max} holds that $\sum_i^{d_{max}} \langle \deg \rangle^i \simeq n$, from which, assuming $\langle \deg \rangle \gg 1$, we solve that $d_{max} \simeq \frac{\log n}{\log \langle \deg \rangle}$. This property was first studied in the context of social networks by de Sola Pool and Kochen [78], and inspired the famous "six degrees of separation" studies by Milgram [271, 272]. It has also been empirically tested on the WWW [6, 51] and the Facebook social network [14]. However, the clustering coefficient in the random model, is inversely proportional to network size, as can be seen below, which underestimates the coefficient values for most real-world networks [22].

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N(i)\}|}{\deg_i(\deg_i - 1)} \simeq p \simeq \frac{\langle \deg \rangle}{n}$$

To compute the degree distribution, we observe that for a vertex of degree k, each of its k edges can be modeled as an independent event of connection to any of the remaining $(n-1)$ vertices of the graph with probability p. Thus, the probability p_k of getting a degree$-k$ vertex follows a Binomial distribution [41], which, for large values of n and small values of p, conditions met by many empirical networks with sparse connectivity, can be approximated

by the Poisson distribution [97],

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \simeq \exp(-\langle \deg \rangle) \frac{\langle \deg \rangle^k}{k!}.$$

The Poisson distribution, in turn, is a discrete probability distribution with a peak around $\lfloor \langle \deg \rangle \rfloor$. However, an important flaw of this distribution is the fact that it underestimates the presence of hubs, or highly connected nodes. This fact became more noticeable at late 90's with the availability and study of network data that revealed that the distribution of several real world networks such as the scientific paper citations [242], the World Wide Web [6, 165], the Internet [96] and metabolic networks [149], can be better approximated by power-law distributions. This distribution is sometimes called *heavy tailed* or *fat tailed* [18], meaning that the decay of its complementary cumulative distribution function (CCDF) at large values, is slower than exponential. For heavy tailed distributions -such as the *log-normal* or *Weibull*-, therefore, it is not rare to encounter very large values or *outliers*. The Poisson distribution, in contrary, is a *thin tailed* distribution, meaning that its decay is exponential. Hence, the outliers are rare, and most drawn values are near to the mean.

In the next subsections, we will present some definitions regarding the power-law distribution.

2.2.3 Power-law distribution

A random variable follows a power-law when its probability density function is given by an expression of the form $p(x) = Kx^{-\alpha}$ where α is called the *exponent* of the power-law, and K is a *normalization constant* depending on α. Few real-world distributions follow a power-law on their whole range; many times the power-law behavior is observed only for higher values, in whose case it is said that the distribution has a *power-law tail*. In analytic terms, since this density function diverges when x goes to 0, there has to be a lower limit $x_{\min}$ from which the power-law holds. That is, $x_{\min}$ is a value that satisfies

$$\int_{x_{\min}}^{\infty} Kx^{-\alpha} dx = 1. \tag{2.1}$$

Moreover, from (2.1) it is simple to solve an expression for K

$$K = (\alpha - 1)(x_{\min})^{\alpha - 1}. \tag{2.2}$$

If we take the logarithm in both sides of the expression defining the probability density of a power-law distribution, we arrive at the expression

$$\ln(p(x)) = -\alpha \ln x + \ln K,$$

which shows the reason that the prototypical representation of a power-law PDF (probability density function) is a straight line on a log-log plot, with slope equal to $-\alpha$. Another typical representation is that given by the complementary empirical cumulative density function (complementary ECDF or empirical CCDF), but in this case the exponent equals $\gamma = \alpha - 1$,

with α being the PDF exponent [206]. In spite of this simple computation, if we need to estimate the exponent of a power-law from a set of observations, the intuitive idea of doing a linear regression on a log-log plot has been reported to be biased and inaccurate [114]. This should be taken as a caution note for future works, since it is a common practice in the literature to appeal to this method to estimate the exponent of the distribution.

2.2.4 Scale-free distribution

Frequently power-law distributions (networks) are also referred to as *Scale-free distributions* (networks). The reason for this is because the power-law distribution is equivalent to a Scale-free distribution. In what follows we will review an argument to give more precision to this claim. Assume there is a probability distribution $p(x)$ that satisfies the Scale-free property, that is,

$$p(bx) = g(b)p(x), \forall b. \tag{2.3}$$

This relation means that if the units or the scale in which x is measured, are increased by a factor b, the shape of the distribution $p(x)$ remains the same, except by a multiplicative constant. The distribution is the same whatever scale we use to record the values. This implies that there is not a typical value for this distribution, nor a proper scale to measure the values of the distribution. These values can span several orders of magnitude, and show a high heterogeneity. Small to moderate values may appear, but also high or even huge values.

Now, from Equation (2.3), we can set $x = 1$, from which we get that $p(b)/p(1) = g(b)$. Then, Equation (2.3) can be rewritten as

$$p(bx) = \frac{p(b)}{p(1)}p(x), \forall b. \tag{2.4}$$

By deriving both sides with respect to b, (2.4) leads to

$$xp'(bx) = \frac{p'(b)}{p(1)}p(x), \forall b, \tag{2.5}$$

where p' corresponds to the derivative of p with respect to its argument. Let us now set $b = 1$ in (2.5) and reorganize some terms to write

$$\frac{dp}{p} = \frac{p'(1)}{p(1)}\frac{dx}{x}.$$

This is a first order differential equation whose solution is given by

$$\ln(p(x)) = \frac{p'(1)}{p(1)}\ln(x) + \ln(p(1)). \tag{2.6}$$

Finally, taking the exponential at both sides we arrive to

$$p(x) = p(1)x^{-\alpha},$$

where $\alpha = -p'(1)/p(1)$, and we verify that $p(x)$ necessarily is a power-law distribution [206].

We close this section with a note about the hypothetical physical origin of these distributions in nature. According to Newman (2005) [206], among the miscellanea proposals of generative mechanisms of scale-free distributions, the most important, in terms of providing plausible and general mechanisms to account for many of the observed distributions, would be the following two. The first is a model by Yule (1925) [295] inspired by statistics of biological taxa. This idea has been reformulated by many authors [255, 108, 235, 199], and it is equivalent to the preferential attachment mechanism of the Barabási-Albert model reviewed in Section 2.3.1. The second one corresponds to the so-called *critical phenomena*, or the associated *Self-organized criticality* [16]. The mechanism in simple terms is the following. Some systems have only a single macroscopic length or size-scale, which, under certain circumstances, or by simple evolution of the system without external perturbation, can diverge, leading to a magnitude without a scale factor to set its size. That is, it is obtained a scale-free or power-law magnitude.

2.2.5 Moments of the power-law

Let x be a power-law distributed variable, whose density is $p(x) = Kx^{-\alpha}$ with lower limit x_{min}. The $\ell-$th moment of x is given by

$$\begin{aligned}\langle x^{\ell}\rangle = \int_{x_{min}}^{\infty} y^{\ell} p(y)dy &= K\int_{x_{min}}^{\infty} y^{\ell-\alpha}dy \\ &= \frac{K}{\ell-\alpha+1}\left(y^{\ell-\alpha+1}\right)_{x_{min}}^{\infty}.\end{aligned}$$

This expression is well defined for $\ell < \alpha - 1$, and in this case, by (2.2), its value is given by

$$\langle x^{\ell}\rangle = \frac{\alpha-1}{\alpha-1-\ell}x_{min}^{\ell}.$$

In particular, the second moment $\langle x^2\rangle$ diverges when the exponent $\alpha \leq 3$, and the first moment, the mean, diverges for exponents $\alpha \leq 2$. The fact that, for instance, the mean is not defined, implies that the mean of finite samples of the power-law variable can exhibit large fluctuations [206]. This is not a purely theoretical consideration, since the exponent values for many real-world distributions are comprised around these values, see Table 2.1. Networks with exponent less than 2, for example, are networks where the number of links grows faster than the number of nodes, a feature that is well represented by the network of software packages [254]. This behavior of the moments of the distribution is essential to analytically prove, for example, the empirical robustness of the Internet under random breakdown of its nodes, or in other words, that a connected cluster survives even for arbitrarily large fractions of crashed sites. This argument may be reviewed in the Annex A.

Network	Type	n	α
film actors	undirected	449913	2.3
telephone call graph	undirected	47000000	2.1
email messages	directed	59912	1.5/2.0
WWW nd.edu	directed	269504	2.1/2.4
WWW Altavista	directed	203549046	2.1/2.7
word co-occurence	undirected	460902	2.7
Internet	undirected	10697	2.5
metabolic network	undirected	765	2.2
protein interactions	undirected	2115	2.4

Table 2.1: **Size and exponents for a number of published networks.** In/out-degree exponents are given for directed graphs. The table is extracted from Newman (2003) [210], where there is a column for the original references where the networks were published.

2.3 Contemporary network science models

This section makes a brief review of the contemporary field of networks science. It starts with some of the first and most studied models of preferential attachment (PA) and other mechanisms that, in some of its variants, turn out to be equivalent to PA. Next sections deal with distinct criticisms to these models, ranging from models merit to recover empirical features, conceptual formulation and fitting to real-world networks. It follows with a bridge section of the general PA growth, which motivated the study of important networks in the field like citation, collaboration, and Internet. The study of citation networks and the hierarchical structure of networks, inspired the development of some later models reviewed in Section 2.3.5 which do simultaneously recover all of the complex networks patterns that the first models do not. The final section introduces a nuance to these unifying aspirations, at least for the study of social networks, in the sense that it reviews converging evidence from distinct works pointing to the idea that social networks would have a weak scale-free character.

2.3.1 First models

Preferential Attachment. "Emergence of scaling in random networks" (1999) [20], written by Albert-László Barabási and Réka Albert, proposed a simple model to explain the power-law degree distribution observed in many real world systems modeled as networks. The authors postulate a random network model based on two principles: growth and preferential attachment. The first refers to the intuition that most real networks are open, formed by the continuous addition of new vertices. This idea is incorporated as follows: the model starts with a small number of vertices m_0, and at every discrete time step t a new vertex is added to the network with $m \leq m_0$ edges connecting the new vertex to vertices already present at time t. The second ingredient is the notion that popular elements of the network are more likely to attract the incoming new elements. This is implemented assuming that the probability that a new vertex v is connected to vertex i, is proportional to the degree $\deg(i,t)$ of that vertex at time t, that is, if our network is the undirected graph $G = (V, E)$,

then

$$\mathbb{P}\left(v \text{ points to } i\right) = \frac{\deg(i,t)}{\sum_j \deg(j,t)}. \tag{2.7}$$

The stationary distribution of the degree of this model is shown to be a power-law of exponent equal to 3. In spite of being developed to capture only one of the patterns of large complex networks, this model has been very influential and is considered one of the seminal works in the contemporary science of complex networks [62]. It provided simple local random rules that converge to a stationary distribution, showing that several complex systems can self-organize to display global features. The idea postulated by this model is, however, not new, and the Barabási-Albert (BA) graphs can be considered the updated network version of this idea. The origin can be traced back to the aforementioned (Section 2.2.4) work by Yule (1925) [295], which proposed an explanation for the distribution of the number of species in a genus, family or other taxonomic group. In his model, the *Yule process*, the probability that a new species is added to a particular genus i having k_i species is proportional to k_i. This idea can be found also in later works with the names of *Proportional growth* [108], *Cumulative advantage* [235], or the *Matthew effect* [199]. In particular, in Price (1976) [235] there is a model for the number of citations received by scientific papers. In this model, new vertices are added over time, and edges starting from these vertices connect to an existing vertex with a probability given by

$$\mathbb{P}\left(\text{to connect new node to existing } v\right) = \frac{k(v) + k_0}{\sum_j (k(j) + k_0)}, \tag{2.8}$$

where $k(j)$ is the current in-degree of node j, and k_0 is a constant put to generate edges at the beginning of the process when all the nodes have in-degree equal to 0. Later models have generalized the preferential attachment model in several ways [84, 25, 85]. An interesting line of models, searches for local mechanisms engendering preferential attachment or power-law degree [277, 263, 83].

Copying models. The work by Kleinberg et al (1999) [165], proposed to conceptualize the World-Wide Web as a directed graph. The authors start by reviewing some algorithms that have been applied to the Web graph: the HITS method [164] and the enumeration of certain bipartite cliques [174]. The first is a method to find relevant pages for a search query, while the second is an algorithm to enumerate cyber-communities, represented as complete bipartite cliques in the graph. Then, they report the power-law distributions of the in-degree and out-degree, and the distributions of the above cyber-communities, observing that usual random graph models cannot account for these measurements. Motivated by previous considerations, the authors present a new random graph model based on the intuition that most page creators will link to pages within some topic, by looking for a list of the topic of interest, and including or *copying* many links from the list to their own page. The description of this copying process is as follows:

1. A node is created at every step.
2. To throw a coin with probability β of obtaining head. If head, it points to a node chosen uniformly at random. Otherwise, it copies a (uniform) randomly sampled edge. (2.9)

The authors inform that simulations suggest that the stationary density function of the degree is a power-law. Later works have studied this mechanism in more depth [173]. The appeal of the Copying model is likely based in its intuitive abstraction of possibly several forms of network evolution. Not just the "copy-and-paste" performed by web creators. A similar mechanism of gene duplication has been proposed as a generative process for protein interaction networks, that can account for the observed power-law degree distributions [278]. However, mathematically it is not so distinct from preferential attachment. We can write the probability $\mathbb{P}(k)$ that the new node is connected to a node of degree k as follows [210]. Suppose we have an undirected graph $G = (V, E)$ of size $N = |V|$. The probability of choosing uniformly a node is $1/N$. The sampling of an edge in step 2. is equivalent to pick a node linked to a randomly selected link. Since there are k edges that lead to each vertex of degree k, the probability of choosing a degree-k node is proportional to k. And since in an undirected network it holds that $\sum_{i \in V} \deg(i) = 2 \cdot |E|$, it follows that

$$\mathbb{P}(k) = \frac{\beta}{N} + \frac{(1-\beta) \cdot k \cdot |\{i \in V | \deg(i) = k\}|}{2 \cdot |E|}. \tag{2.10}$$

Equation (2.10) shows that $\mathbb{P}(k)$ is proportional to k, analogously to the BA model.

Optimizing processes. The paper by Ferrer i Cancho and Solé (2003) [57], offers an explanation about the origin of power-law networks as an outcome of a network global optimization process. This model is motivated by, among others, the role of network optimization in explaining allometric scaling [287] or the argument that metabolic pathways would have been optimized [57]. The model starts on time $t = 0$ with a random graph in which two nodes are connected with probability p. The number of nodes remain fixed along the process. The objective function to minimize is a linear combination of the cost of physical links, and the communication speed, or in other words,

$$E(\lambda, G_t) = \lambda d(G_t) + (1-\lambda)\rho(G_t),$$

where G_t is the network at time $t \geq 0$, $0 \leq \rho(G_t) \leq 1$ is the normalized number of edges of the network G_t, and $0 \leq d(G_t) \leq 1$ is the normalized average vertex-vertex distance on G_t. The minimization proceeds as follows: at time $t > 0$, each possible edge of G_t can be switched with probability μ, that is, added if it is not present, or deleted if it is not. The new network G_{t+1} is accepted if $E(\lambda, G_{t+1}) < E(\lambda, G_t)$. Otherwise, another set of changes is performed. The algorithm stops when the changes are not accepted some M number of times consecutively. The networks produced in this process can have very different structural features. With λ near to 0, the networks have a random connectivity, while for greater λ, hubs formation explains the transition to a power-law distribution. For still greater values

of λ, hub competition precedes the emergence of a central vertex every vertex is connected to. Finally, when λ approaches 1, a dense graph emerges, by means of a progressive increase in the average degree of non-central vertices and a sudden loss of the central vertex.

Optimization principles have some history in explaining power-law distributions. One of the first works in this line is Mandelbrot [192], in which the power-laws are the solution for the word frequencies of languages that maximize the transmitted information. *Highly optimized tolerance* (HOT) [58] is a more recent model that asserts that power-laws arise in designed, highly structured systems due to trade-offs between yield, cost of resources, and tolerance to uncertain risks. A later model inspired in the trade-offs idea from HOT model, is the *Heuristically optimized trade-offs* [95]. This is a network growth model that shows that power-law degree distributions can result from local optimization, or optimization performed by each new arriving node, of both "last mile" connection costs, in that it is easier to connect to a nearer node, and transmission delays, or the goal of connecting to central nodes. Finally, Berger et al (2004) [31] shows that these growth processes with competing trade-offs forces may be equivalent to a generalized form of preferential attachment.

2.3.2 Conceptual critiques to the first models

The field of complex networks has the appeal of showing, by means of simple models, that many systems coming from diverse sources would have a sort of universal patterns, independent of the subtleties of every domain. However, the "scale-free story" has not been exempt from criticisms. Most of these are directed to the preferential attachment principle in its original form, probably due to the fact it is the most studied in the field. We briefly summarize here some of the main conceptual objections in the literature. Next subsection addresses objections related to the difficulties in determining whether a real-world network is scale-free.

Rigorous definition. Since the results in the BA model were derived with a continuous approximation for the asymptotic limit of large network sizes, it has been argued that this model lacks a precise mathematical definition. Motivated by this, Bollobás and collaborators (2001) [43] have proposed a reformulation of the Barabási-Albert model, consisting in a random graph process which is more suitable for mathematical treatment. The conclusions drawn from this model [42] are similar to those from the conventional approaches. Another recent proposal in this line is, for example, Ostroumova et al (2013) [216].

Recovered properties. First models reviewed in the last section are not able to recover simultaneously the three properties of complex networks seen in Section 2.2.1, that is, power-law degree distribution, high clustering and small diameter. The random network reviewed in Section 2.2.2 only recovers the small diameter as shown therein. The preferential attachment mechanism displays power-law degree and small diameter of order $\log n / \log \log n$ [42], but its clustering is of order $(\log n)^2/n$ [44], that is, the clustering approaches 0 as the network grows, in contrast to empirical networks. The famous model by Watts and Strogatz (1998) [286], which departs from a regular lattice by randomly rewiring links with a probability

$p \ll 1$, is able to recover the high clustering and small diameter as long as the network is large enough, but the degree distribution is centered around the mean value [23]. These negative modeling results fueled many of the later modeling efforts reviewed in Section 2.3.5.

Characterization by means of the degree distribution. Li, Alderson, Doyle and Willinger (2005) [184] claim that the scale-free condition of a network cannot be exclusively given by its power-law degree, because of the enormous diversity of graphs associated to the same degree sequence. As an indicative example of this, the authors show that two networks with the same power-law degree sequence, one generated from the hierarchical model by Ravasz et al (2002) [239], and the other with a HOT model [185], exhibit a completely different behavior when extracting the highest-degree nodes. The first becomes disconnected, while the second gets only a minor disruption. To account for this result, the authors introduce the s measure, which quantifies the extent to which a graph has a hub-like core, and is maximized when high-degree nodes are connected to other high-degree nodes. The authors then report that the BA model generates graphs with high s values, the same as the hierarchical model, and in contrast with the low values of the HOT model. These observations, together to other results, make the authors claim that the conventional scale-free graphs represented with the BA model, can be re-defined as graphs with a power-law degree sequence and a high value of s, recovering thus the known properties of these models. Finally, the authors discuss that the Internet router-level topology is qualitatively more similar to the HOT models, and that the "robust, yet fragile" claims of the Internet are wrong and based mainly on ambiguous data. A similar observation was made by Tanaka (2005) [269], stating that random models, as the BA graph, are incompatible with the highly structured networks resulting from engineering design or biological evolution, an outcome that would be better captured by HOT models [58]. A later work [297] has made a similar argument to that by Li et al (2005) by proposing a maximum entropy process to represent the typical properties of scale-free networks.

Another interesting proposal is due to Zhou et al (2020) [298], which proposes that a new measure based on the links of a network, the *degree-degree distance*, would usually exhibit a stronger power-law than the degree distribution of a finite-size network. This measure, for each link, corresponds to the dimensionless ratio between the maximum and the minimum between the two values of the degree at each link end-point. Comparison to the degree distribution, and statistical fits in several empirical networks, are used by the authors to conclude that this measure would provide a better characterization of a finite-size network being scale-free than the degree distribution. This proposal is given in the context of the recent controversies regarding the determination of the scale-free property on real-world networks, a topic addressed in the next subsection.

2.3.3 Critiques in fitting real-world networks

Determining whether an empirical network is scale-free, is not an easy question. It has been reported that for some scale-free networks the distribution of the entire network, and that of randomly sampled sub-networks may differ [266], or that standard maximum-likelihood approaches lead to false rejections of statistical laws in the presence of correlations in the data [107]. But there is a deeper conceptual issue. The usual definition of the scale-free

degree distribution relies on a continuous approximation valid in the asymptotic limit of large networks. In real simple networks, the degree is always bounded by the number of nodes in the network, which introduces finite-size effects such as the existence of a cutoff or bound for the maximum degree, which may be due to several factors [39]. In terms of the statistical fit, there is always the possibility of a better fit achieved by alternative models for heavy-tailed distributions such as the log-normal [203]. In the next paragraph, it is described the probably most accepted method for statistical inference of power-law distributions in the last decade. The subsequent paragraph describes more recent questioning to the scale-free character of empirical networks, which lead to the development of a recent method which is arguably more suitable for inference on real-world datasets.

Power-Law parameter estimation. A simple and reliable way to estimate the exponent α from a sample $\{x_i\}_{i=1}^m$ of m observations from a power-law distribution is to employ the maximum likelihood estimator (MLE) given by the formula [206],

$$\hat{\alpha} = 1 + m \left(\sum_{i=1}^{m} \ln \frac{x_i}{x_{\min}} \right)^{-1}. \tag{2.11}$$

As it is apparent from the previous formula, the quality of the estimate $\hat{\alpha}$ depends on the estimate $\hat{x}_{\min}$ of the lower limit $x_{\min}$ of the distribution. A possible method to estimate this, is the one proposed by Clauset et al (2007) [68, 67], which selects the $\hat{x}_{\min}$ that makes the distributions of the empirical data and its fitted power-law model as similar as possible above $\hat{x}_{\min}$, that is, where the fit model is well defined. This similarity, or equivalently, distance between two probability distributions, could be implemented through the Kolmogorov-Smirnov statistic (KS), whose expression in this case is the following

$$D(x_{min}) = \max_{x \geq x_{\min}} |S(x) - P(x)|, \tag{2.12}$$

where $S(x)$ is the CDF of the data for observations with value greater than $x_{\min}$, and $P(x)$ is the CDF of the power-law model that best fits the data (for example, the MLE estimation (2.11)) in the region $x \geq x_{\min}$. Thus, $\hat{x}_{\min}$ is chosen as the value $x_{\min}$ that minimizes $D(x_{\min})$.

Recent controversy. Broido and Clauset (2019) [53] used the methodology in the last paragraph to propose several criteria of being scale-free with increasing evidence strength, to then analyze a big corpus of nearly 1000 empirical networks. The authors conclude that strong scale-free structure is empirically rare, and that for most networks, log-normal distributions fit the data as well or better than power laws. This work generated a lot of subsequent discussion and new works [19, 298, 281]. For example, it has been argued that the same work reported that more than half of the dataset is consistent with a power-law fit with cutoff, or that the proposed tests do not work well with the controls given by some theoretical models [19]. An important point was made by Voitalov et al (2019) [281], where a new definition of power-law distribution is proposed, equivalent to regularly varying distributions allowing deviations from pure power-laws arbitrarily, but without affecting the power-law tail exponent. Thus, the method would be more suitable in dealing with real datasets, and according to the article scale-free networks would not be rare since, for example, in a dataset of 115 real-world networks, 49% of the considered undirected networks have degree sequences that are power-law under this definition.

2.3.4 Preferential attachment growth

In this section we review an early generalization of the preferential attachment hypothesis, that will serve as a bridge to the next sections addressing posterior models of complex networks. The preferential attachment hypothesis states that the rate $\Pi(k)$ in which a $k-$degree node creates new links is an increasing linear function of k, and though Barabási and Albert (1999) suggested this rate may follow a power-law, initial evidence from simulations showed that the scale-free property emerged only in the linear case [20]. This intuition was confirmed by Krapivsky et al (2000) [171], where the following was found. Suppose the rate $\Pi(k)$ may have the following general form,

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha} = C(t)k_i^\alpha.$$

Krapivsky et al (2000) prove that for $\alpha = 1$ this model reduces to the usual BA graph with exponent 3. The original formulation by Krapivsky et al (2000) in fact is more general and allows tuning the exponent of the power-law degree distribution to every value larger than 2. In the sublinear case, $\alpha < 1$, the degree distribution follows a stretched exponential, that is, the bias favoring more connected nodes is weaker, which produces an exponential cutoff that limits the size of hubs. On the other hand, for a superlinear attachment $\alpha > 1$, a single node becomes central and connects to nearly all other nodes.

Posterior works have studied dynamical data on real evolving networks to test the adequacy of this model. Jeong et al (2003) [148] found that Internet and citation networks present linear attachment, while scientific collaboration and the Hollywood actor networks are consistent with sublinear attachment. Pastor-Satorras et al (2001) [220] also finds linear attachment in the Internet network, Newman (2001) [208] reports that scientific collaborations are consistent with linear attachment, in contrast to Barabási et al (2002) [21], which finds that these networks are best described by sublinear scaling. Redner (2004) [243] also finds linear attachment in citation networks. The evolving networks studied in these works are important for the development of better models in network science, hence, in the next section we focus on later models inspired by intuitions of social dynamics coming mainly from citation networks.

We do not address in this document the other types of networks that likely present linear attachment and scale-free degree, the scientific collaboration and the Internet networks. The former kind of networks comprise very complicated systems where there are at least two distinct entities, the authors and their collaborations (papers), which is the reason for the appearance of increasingly complex elements such as new links among existing nodes [21], weighted arcs [186], *affiliation* bipartite graphs, where there are sets of paper nodes and author nodes [182], or the fact that the betweenness centrality would be a better driver for preferential attachment than the degree [1]. For simplicity we do not address them, though some models for scientific collaboration have delivered interesting insights [161, 186]. In the case of the Internet, we do not address it since other mechanisms that this survey do not cover such as an underlying hyperbolic geometry [172, 37], have been proposed as generative models.

2.3.5 Posterior models of complex networks

We review now a sample of models proposed during the last two decades, that contributed original mechanisms to simultaneously recover the properties that preferential attachment does not: the scale-free and the small-world properties, as it is discussed in Section 2.3.2. We group these models in two topics that are relevant for advancement in knowledge of social dynamics: the citation networks, and the hierarchical structure of networks. Citation networks have nodes representing documents, usually published papers, and the directed links represent the citations received. These networks, as said in the last subsection, present linear preferential attachment, and though they may be classified more generally as information networks rather than strict social networks [210], their study has revealed interesting insights related to the formation of triads and clustering. This idea is consistent with the empirical finding that, in the formation of new links in information networks such as citation and the web, local search through the current network has a more important weight than in more purely social networks, where uniformly random meetings are more prevalent [143]. Citation networks have an exponential out-degree distribution with a maximum at intermediate values [276], and considering that the out-degree of a node does not change once created, it is more interesting to study the in-degree distribution which follows a power-law [243]. These networks also present complex interactions between age and citations, which have raised questioning to the scale-free character of them [115].

The second topic consists in the study of the mesoscopic structure of networks, that is, an intermediate level between the local and the global, which comprises the possible existence of communities and their inter-connectivity. This structure may be key to explain, for example, different cooperation levels in evolutionary games, that may not be understood only in terms of global network properties [190]. The community structure is a hard problem in itself, from its conceptual definition, detection and combinatorial enumeration [103], to the fact that communities may overlap and display scaling properties [217]. In this subsection, we will make a brief review on some foundamental insights relating this structure to some network properties such as the clustering.

Citation networks

Holme and Kim (2002) [135] extend the standard BA network model to include a "triad formation step". That is, if an edge between the new node v and w was added in the previous preferential attachment step, then with certain probability add one more edge from v to a randomly chosen neighbor of w. This step is inspired by the sociological mechanism of "sibling bias" [104], where after being acquainted with (linked to) w, an actor v is likely to be acquainted to w's acquaintances as well. Thus, this model possesses the same characteristics as the standard scale-free networks like the power-law degree distribution and the small average geodesic length, but with high-clustering at the same time. A later model [73] based on this work is able to tune the values of the power-law exponent and the average local clustering.

Klemm and Eguíluz (2002) [166] observe that there is a discrepancy between the BA model and empirical networks, in the correlation between a node's age and its rate of acquiring links.

For the network of scientific citations this correlation is negative: the mean rate of citations a paper receives decreases with increasing age. On the other hand, in the BA model the attachment rate is proportional to the degree, being largest for the oldest nodes, producing a positive correlation between the mean attachment rate and age. This modeling defect motivates a new principle based on degree-dependent deactivation of nodes, the mechanism being as follows. Each node of the network can be in two different states: a new node added starts in active state, and later it randomly deactivates, and remains inactive. A node only receives links from subsequent nodes in the active state, and the probability of deactivation decreases with the in-degree of the node. The transition of a node from the active to the inactive state can be interpreted as a collective "forgetting" of the node since new nodes do not connect to it anymore. The model can recover the power-law distribution of degree, linear preferential attachment of new links and a negative correlation between the age of a node and its link attachment rate. Additionally, the clustering reaches an asymptotic value larger than for regular lattices of the same average connectivity.

Leskovec et al (2005) [182], is a large empirical and analytical study about some patterns that emerge along the evolution of networks on time. The paper takes snapshots of several networks at regularly spaced time points, among which there are several citation graphs, Internet routers, and affiliation graphs. The study reports that, on every of these datasets, there is *densification* of graphs, with the number of edges growing super-linearly over time in the number of nodes, or in other words,

$$e(t) \propto n(t)^a,$$

where $e(t), n(t)$ are the edges and nodes of the graph at time t, and a is an exponent found to have a value between 1 and 2. The other phenomenon reported is that of the *shrinking diameter*, a decreasing trend of the diameter over time, seemingly converging to some asymptotic value, observed for all the graphs. These findings challenge the usual assumptions of the constant average degree over time, and the diameter as a slowly growing function of network size [182]. To account for these patterns, several generative network models with increasing complexity are proposed, which exhibit as emergent properties the reported empirical behaviors. In particular, the *Forest Fire* model displays power-law distributed in- and out-degrees, densification and shrinking diameter. This model specifies a probabilistic dynamics of a "burning" of nodes, which spreads to its neighbors and then recursively until it dies out. The mechanism is inspired by the dynamics of creation of paper citations: an author of a new paper consults a reference, follows some of its references, and then continues accumulating references recursively. It may be considered a generalization of triad formation. In Ren et al (2012) [244] this model is contrasted against empirical data to propose a new model where triad formation to some node j's neighbor is extended to all nodes in the clique where j belongs to, in order to reproduce the number of triangles, the high clustering coefficient and the size distribution of co-citation clusters in empirical networks.

Hierarchical structure

Ravasz and Barabási (2003) [238] present a deterministic network model having both power-law degree distribution and average clustering coefficient approximately equal to 0.743. The model follows an iterative process, generating a hierarchical organization in which small

degree nodes belong to dense communities, and hence have high clustering coefficients, and on the other hand, the hubs connect different communities, hence resulting in low clustering. This hierarchical organization is expressed in the following expression, which is referred to as the *clustering hierarchy*,

$$C(k) \sim k^{-1}, \tag{2.13}$$

where k is the degree and $C(k)$ is the average of local clustering among $k-$degree nodes. The results show, by numerical simulations, that not only the deterministic model satisfy relation (2.13), but it also holds on several empirical networks (actors, language, WWW, Internet at the autonomous level). The authors report two real networks that do not follow (2.13), the internet at the router level and the power grid, suggesting that their geographical nature might impose strong restrictions to the hierarchical topology. Later works [264, 140] have refined (2.13), for example, by arguing that the exponent is more general and depends on the power-law of the degree distribution, and that for low values of k, $C(k)$ has a flat region [264].

Xuan et al (2006) [294] propose a model where there is a growing number of vertices, but a fixed number of hierarchical, modular levels. The intuition is to capture social systems, large-scale logistic systems, or intercontinental air transportation networks, for example, where there are fixed hierarchical and modular structures mainly determined by geographic factors, but the vertices of them increase very quickly. The model has a treelike structure: at the lowest level there are several groups of vertices, forming modules of this level. Those modules are in turn grouped to constitute higher-level modules and so on. There are two stochastic events of network growth: in-module connection and between-module connection. The creation of connections in each case obeys preferential attachment. The results show that the degree distribution, the module size distribution, and the clustering function of the model possess a power-law property which is similar to that in many real-world networks.

This idea of the influence of the mesoscopic structure to determine the local connections is reinforced by Tomasello et al (2014) [270]. This work studies the growth of R&D networks by proposing a network model where nodes are firms, and links their alliances. Nodes are endowed with the attributes of activity and label. The activity is the propensity to be involved in an R&D alliance event, which encodes the instantaneous network dynamics and provides an explanation for the presence of hubs [228]. The label represents the belonging of a firm to a group, and is propagated by alliance creation events. By fitting only three macroscopic properties of the network topology, this framework is able to reproduce a number of micro-level measures, including the distributions of degree, local clustering, path length and component size and the emergence of network clusters. By estimating the link probabilities towards newcomers and established firms from the data, the authors find that endogenous mechanisms (previous alliances and previous network structures) are predominant over the exogenous ones (exploratory search of new partners) in the network formation, even considering that in this model the groups result from self-organizing processes, rather than being imposed as in the previous reviewed work.

An important point was made by Colomer-de-Simón et al (2013) [74], where it is asked whether triangles in real-world networks are nearer to maximally random graphs with prescribed degree and clustering distributions [219], or to maximally ordered graphs where triangles are forced into modules by specifying the joint degree-and-clique-size distribution [113].

The approach was to study *m-core* or *k-dense* landscapes [119], where the m-core is the maximal subgraph with edges participating in at least m triangles. The intuition of this measure is that each edge belonging to a clique of size n participates in $(n-2)$ distinct triangles, whereas an edge connecting two cliques participates in 0 triangles. This property defines a set of nested subgraphs that is able to distinguish between hierarchical, where m-cores do not fragment, and modular architectures, where m-cores are always fragmented. The study finds that the clustering organization in real networks is neither completely random nor ordered although it is more random than modular, that is, there is a strong hierarchical structure, where each layer of nodes with the same m-coreness is nested within the previous layer, which is better reproduced by maximally random graphs. This suggests that the structure of real networks may in fact be the outcome of self-organized processes based on local optimization rules, in contrast to global optimization principles.

2.3.6 Social networks are only weakly scale-free

In contrast to the models reviewed in the last subsection, which are able to recover either the scale-free and the small-world properties, we will now review evidence from distinct works pointing to the idea that social networks would have a weak scale-free character. In this way, this subsection introduces a nuance to those modeling efforts reviewed in the previous subsection, in the sense that complex networks patterns would have different importance and interplay depending on the specific type of network studied. For this reason, if the goal is to study social networks or network societies, it is desirable to bear in mind the explicit social mechanism under study, which is the subject of Section 2.4. The present section either introduces new works, and summarizes some previously discussed studies.

We have already reviewed in Section 2.3.4 several papers which, by studying dynamical data of empirical networks with timestamped node connection events, looked for measuring the growth rate of preferential attachment. In general, these works find that Internet, or citation networks present linear attachment or scale-free degree [220, 243, 148], while more social networks like scientific collaborations or the Hollywood actors present sublinear attachment or exponential degree distribution [148, 21]. The hypothesis of the weak scale-free character of social networks is more formally tested by Jackson and Rogers (2007) [143], where a dynamic model of network formation is proposed in which nodes find other nodes to form links in two ways: some are found uniformly at random, while others are found by searching locally through the current structure of the network, like meeting friends of friends. The model is fitted to six network datasets, finding that the relative roles ratio of the random versus network-based meetings is more than seven times greater in a co-authorship network than in a World Wide Web or citation network, and that the formation process is almost uniformly random in two networks where the links correspond to friendships. The result that an effective preferential attachment is the outcome of growing network models based on local rules such as, for example, a random walk that adds new links to neighbors of connected nodes, was studied previously by Vázquez (2003) [277]. Finally, Broido and Clauset (2019) [53], who put a test on the scale-free distribution in a wide corpus of diverse empirical network datasets, show that half of the social networks studied lack any direct or indirect evidence of scale-free structure, and that most of the networks, including scientific collaboration or board of directors, fall into the weakest evidence categories.

2.4 Models of communities in networks

This section reviews a series of diverse studies which, by means of proposing a model of some specific social dynamics in networks, attain to recover as a result the presence of communities, or cohesive groups. These works may be roughly classified in terms of the methodology they employ. Models coming mostly from the physics literature work with stochastic processes of network formation. On the other hand, models from the economics literature use game-theoretic concepts to describe strategic processes of network formation.

However, since we are more interested in the hypotheses formalized, the kind of social behaviors captured by each model, we will structure this section based upon this criterion. Most of these ideas are inspired by theories of trust and social capital [71, 117] that claim that closure, or the reinforcement of relationships by the presence of common friends, promotes cooperation. Though these accounts have been supported by experiments and field studies [112, 196, 274], they have only recently been formulated as formal models by some of the studies surveyed in this section. This kind of approach allows testing whether precise structural patterns of social networks are the outcome, or are optimally arranged for a given behavior. Thus, the precise notion of community varies depending on the work. Main mechanisms modeled can be grouped into complementarity, social sanctioning, favor exchange or risk-sharing, and homophily. We next briefly review each of these themes.

2.4.1 Complementarity

We say that a behavior presents high level of complementarity when an individual requires that many of his or her friends adopt that behavior before being willing to do so. An empirical example of this phenomenon is provided by Centola (2010) [59], which shows that individual adoption of a health behavior is much more likely when participants received social reinforcement from multiple neighbors in an Internet-based social network. A simple model that recovers this intuition is that by Chwe (1999) [66], where each agent in a group wants to participate in a collective action only if the total number taking part is at least her idiosyncratic threshold value. An agent knows the thresholds and links of only the people in her neighborhood in a network of social connections. One result of the model is that low-threshold-agents are much more likely to participate if they are part of tightly knit, high clustering groups of friends. These facts are consistent with another insight in network science posing that small-world networks are highly efficient for information exchange [178]. Finally, a model of static games with linear best replies on networks by Bramoullé et al (2014) [46] gives another consistent intuition for the converse case when the agents' actions are strategic substitutes, that is, situations where actions by other agents may inhibit an agent's own action. This situation comprises, for example, private provision of public goods and oligopolies. The model results show that for a greater degree of substitutability, the equilibrium agents' contributions may be concentrated on some nodes in the network, while for example, more dense local communities may lead to a better diffusion of agents' actions.

2.4.2 Social sanctioning

This kind of studies [17, 126, 187] aim to formalize social mechanisms of misconduct deterrence to foster cooperation such as defecting as punishment on past defections, or ostracism, by analyzing equilibrium outcomes in repeated games played in networks, where the network may provide the structure of either local monitoring of histories of play, or game interactions. For example, Lippert and Spagnolo (2011) [187] proposes an asymmetric Prisoner's Dilemma game where agents play repeatedly in pairs with those agents linked by a network structure. After each game iteration they have the option of passing on observed or received private information on the history of play, and whether to do it truthfully or lie. The analysis shows that cooperating agents do have incentives to transmit information truthfully until the initial deviator is punished, making relations sustainable that would not be so bilaterally (ie, without the network structure). On the other hand, network structures such as trees or stars are never sustainable because agents with only an outgoing link cannot be sanctioned if they defect, in contrary to networks having cycles which are sustainable by ensuring that all defections can be met with punishment. This result provides an intuitive explanation for the importance of closure and density of social networks in social capital theories.

As another example, Balmaceda and Escobar (2017) [17] studies a trust game where there are N investors, and at each round one investor is randomly selected to play with a special non-investor agent. In this game, the investor decides whether or not to participate, and in the first case he chooses an investment level, where then the agent chooses whether to cooperate or to defect. There is also a social network determining the information each investor has. One of the main results is that when a defection is observed only by the victim's connections, cohesive networks (that is, with complete components) are Pareto efficient as they allow players to coordinate their punishments to attain high equilibrium payoffs, decreasing the agent's temptation to defect.

2.4.3 Risk-sharing

This principle refers to the networks representing income transfers functioning as insurance against income fluctuations. Several of these works are inspired by informal lending studied in developing countries where many times these systems replace formal legal institutions of credit [9, 156, 142, 36]. We review next, as a representative example, the idea proposed in the work by Mobius and colleagues [9, 156], which aims to model local transfers between socially close households that take the form of loans. Since the model assumes that agents derive utility from the strength of their relationships with friends, when a borrower needs an asset of a lender, the network connections between individuals are used as a social collateral to secure informal borrowing. The mechanism is the loss of the borrower-lender relationship if the promise of a transfer arrangement in case the borrower fails to return the asset, is broken by the borrower. Thus, the model achieves to predict that dense networks (in terms of the relation between number of paths and reachable nodes) generate bonding social capital that allows transacting valuable assets, while loose networks create bridging social capital that improves access to cheap favors such as information.

A more abstract model is the one by Vega-Redondo (2006) [279], where agents play repeat-

edly an idiosyncratic Prisoner's Dilemma with their neighbors. The social network specifies not only the playing partners but also determines how new cooperation opportunities are found. Search of new links plays a key role since the environment is subject to payoff volatility. One of the model's results is that as volatility rises, the network endogenously increases its cohesiveness (average neighbor distance) due to an enhanced deterrence of opportunistic behavior by neighbors' punishment. However, instead of several communities, there is a large component including most of the connected individuals. This shows the particular relevance of social sanctioning in risky environments.

2.4.4 Homophily

Homophily is the observed pervasive tendency [198] that similar individuals tend to interact at a higher rate than dissimilar people. This pattern is the result both of individual preferences for similarity, and the structure of meetings given by homogeneous group composition [77, 197]. It has been formally argued that preferences for similarity may be the result of limits on the number of relationships an individual can mantain, and direct externalities of investments on a link given, for example, by reciprocity of friendships [247, 45]. On the other hand, homophily can have important effects both in enhancing or impeding information diffusion [116], which may impact adoption of health behavior [60], mate selection [256] or employment opportunities [27].

Regarding its role as a factor for community structure in formal models, it is usually posed as an explanation for culturally homogeneous groups and global dissimilarity among groups, which also requires the mechanism of social influence or induced homophily, that is, the more people interact, the more similar they become [12, 61, 262]. For example, Starnini et al (2016) [262] shows that agents provided with a mobility scheme reproducing empirical data on human face-to-face interactions, plus homophily implemented as a confirmation bias in opinions and social influence in terms of a local search for consensus, can yield the emergence of a metapopulation structure, where there are physically segregated groups of individuals sharing similar opinions. This kind of models have been used to explain the emergence of polarized opinions [98, 26].

2.5 Concluding remarks

Complex networks have been an enormous advance to our understanding of many complex systems and their common features. These comprise the power-law degree distribution, the high clustering and the small network diameter. The first is related to the unbounded variability of the degree, which in turn explains some particular properties of the distribution, as seen in Section 2.2.5. The next two attributes relate to the efficiency of diffusion processes they provide, see Section 2.4.1, present in many diverse systems, particularly in social networks [272]. Despite many important advancements have been produced with the goal of integrating these properties, see Section 2.3.5, one must proceed with care when studying a particular system. Distinct works reviewed in Section 2.3.6 provide converging evidence that pure social networks, in contrast to some information networks such as citation or the

web, would be weakly scale-free. This result is maybe due to the fact that link creation is more uniformly random in social networks, than guided by the local network structure as in information networks [143], where the latter makes that higher degree nodes are more likely to be found since more paths lead to them, as random walk models show [277]. For this reason, it is key to keep in mind the specific mechanisms driving network dynamics, as those reviewed in Section 2.4 driving some network societies.

Chapter 3

QuickCent: a fast and frugal heuristic for harmonic centrality estimation on scale-free graphs

Abstract

We present a simple and quick method to approximate network centrality indexes. Our approach, called *QuickCent*, is inspired by so-called *fast and frugal* heuristics, which are heuristics initially proposed to model some human decision and inference processes. The centrality index that we estimate is the *harmonic* centrality, which is a measure based on shortest-path distances, so infeasible to compute on large networks. We compare *QuickCent* with known machine learning algorithms on synthetic data generated with preferential attachment, and some empirical networks. Our experiments show that *QuickCent* is able to make estimates that are competitive in accuracy with the best alternative methods tested, either on synthetic scale-free networks or empirical networks. QuickCent has the feature of achieving low error variance estimates, even with a small training set. Moreover, *QuickCent* is comparable in efficiency –accuracy and time cost– to more complex methods. We discuss and provide some insight into how QuickCent exploits the fact that in some networks, such as those generated by preferential attachment, local density measures such as the in-degree, can be a proxy for the size of the network region to which a node has access, opening up the possibility of approximating centrality indices based on size such as the harmonic centrality. Our initial results show that simple heuristics and biologically inspired computational methods are a promising line of research in the context of network measure estimations.

3.1 Introduction

Heuristics are proposed as a model of cognitive processes. Some models based on heuristics have been proposed to account for cognitive mechanisms [273], which assume that, though these heuristics are used at a lesser computational cost, they sacrifice accuracy and lead to systematic errors. This viewpoint has been challenged by the so-called *Fast and frugal* heuristics [109], which are simple heuristics initially proposed to model some human decision and inference processes. They have shown that very simple human-inspired methods, by relying on statistical patterns of the data, can reach accurate results, in some cases even better than methods based on more information or complex computations [159, 109]. Due to these features, fast and frugal heuristics have been applied in problems different from their original motivation, including medical decision-making [13], predicting the outcomes of sport matches [253] and geographic profiling [259].

The problem of centrality computation. In this chapter, we provide an example of the usefulness of one of these simple heuristics for estimating the centrality index in a network. Roughly speaking, the centrality index is a measure of the importance of a node in a network. We chose to estimate the *harmonic centrality* index [193] since it satisfies a set of necessary axioms that any centrality should meet [40], namely that nodes belonging to large groups are important (*size* axiom); that nodes with a denser neighborhood, i.e. with more connections, are more important (*density* axiom); and that the importance increases with the addition of an arc (*score-monotonicity* axiom). Consider a directed graph $G = (V, A)$, with V the set of nodes and A the set of arcs or edges. Formally, let $d_G(y, x)$ be the length of the shortest path from node y to x in the digraph G. The harmonic centrality of x is computed as

$$H_G(x) = \sum_{y \in V, y \neq x} \frac{1}{d_G(y, x)},$$

which has the nice property of managing unreachable nodes in a clean way.

Besides its good properties, to compute the harmonic centrality for all nodes in a network we need first to solve the all-pairs shortest-path problem. Notice that by the total number of pairs of nodes, there is an intrinsic lower bound of $\mid V \mid^2$ for computing this centrality, and $O(\mid V \mid^2)$ is already a huge constraint for modern networks. There has been a lot of work on optimizing the computation of all-pairs shortest-paths for weighted networks [230, 229, 234] but even under strict constraints on the structure of the networks [234] this computation is unfeasible for networks with a large number of nodes, usually needing time $O(\mid A \mid \cdot \mid V \mid)$. Thus, in order to use harmonic centrality in practice we need ways of estimating or approximating it.

Though there are few centrality indexes satisfying the three axioms [40], some simple measures can be built that do satisfy them. One way of doing this, is by taking the simple product of a density measure, such as the in-degree, with a size measure, such as the number of weakly reachable nodes [40]. While the in-degree is cheap to compute, many times stored as an attribute so accessible in constant time, size measures have a higher time complexity. For example, the number of reachable nodes, for each node, can be computed from the condensation digraph of strongly connected components, which may give, in the worst case, a total time complexity of $O(\mid A \mid \cdot \mid V \mid + \mid V \mid^2)$. In this chapter, we explore whether expensive indexes, sensitive to either density and size, such as the harmonic centrality, may be approximated by cheap local density measures such as the in-degree.

Our proposal. Our proposed method, called *QuickCent*, is a modification of QuickEst [132], a heuristic proposed to represent the processes underlying human quantitative estimation. QuickCent can be considered as a generalization of QuickEst, in the sense that, albeit in this work we focus on centrality approximation, it proposes a general procedure to regress a variable on a predictor when some assumptions are met. QuickCent is a very simple heuristic based on sequences of *binary clues* associated with nodes in a network; the value of a clue is an indicator of the presence or absence of an attribute signal of greater centrality for a node. The method simply finds the first clue with value 0 (absence), and it outputs an estimate according to this clue. All the clues used in QuickCent are based on the in-degree of the node, thus QuickCent can be seen as a method to regress a variable (harmonic centrality)

that correlates with a predictor variable (in-degree) that is cheaper to compute. Another key characteristic of QuickCent is that it is designed to estimate magnitudes distributed according to a power-law [206], which can model a wide range of natural and human-made phenomena. This chapter extends previous work by some of the authors, mainly by adding the study of networks defying the heuristics assumptions and the performance over empirical networks [232]. Besides that, some technical implementation details, together with the examples and the insight of why harmonic centrality can be estimated with in-degree, are new.

Results and future work. Our method is able to generate accurate estimates even if trained with a small proportion –10%– of the dataset. We compare QuickCent with three standard machine learning algorithms trained with the same predictor variable over synthetic data and some empirical networks. Our results show that QuickCent is comparable in accuracy to the best-competing methods tested, and has the lowest error variance. Moreover, the time cost of QuickCent is in the middle range compared to the other methods, even though we developed a naive version of QuickCent. We also discuss how QuickCent exploits the fact that in some networks, where higher degree nodes are more likely to be found because more paths lead to them, local density measures such as in-degree can be a good proxy for the size of the network region to which a node has access, opening up the possibility of approximating centrality indices based on size, such as harmonic centrality. This insight supports the conjecture that QuickCent may be better suited to information networks, such as the Internet, citation, or scientific collaboration, which can be well approximated by the preferential attachment growth mechanism [20, 148, 277], than to more purely social networks [143, 53], which is an interesting question for future work. Also, working in the future with more general notions of local density [175, 29] may serve to extend the validity of the heuristics for more general networks. The results of this chapter are a proof of concept to illustrate the potential of using methods based on simple heuristics to estimate some network measures. Whether or not these heuristics provide a realistic model of human cognition, is a wide problem [52] which is out of the scope of this work.

Structure of the chapter. The rest of this chapter is structured as follows. We begin in Section 3.2 by introducing the general mechanism of QuickCent, while Section 3.3 presents our concrete implementation. In Section 3.4, we present the results of our proposal, including the comparison with other machine learning methods on either synthetic or empirical networks. Section 3.5 gives a final discussion of the results including directions for possible future work.

3.2 The QuickCent Heuristic

In this section, we give a general abstract overview of our proposal, which we call QuickCent. The setting for QuickCent is as follows: the input is a network $G = (V, A)$ and we want to get an accurate estimate of the value of a centrality function $f_C : V \longrightarrow \mathbb{R}$. That is, for every $v \in V$, we want to compute a value $\tilde{f}_v$ that is an estimation of $f_C(v)$. In our abstract formulation, it does not matter which particular centrality function we are estimating, and

the details of the implementation of the heuristic for the particular case of harmonic centrality are given in the next section. We next explain the general abstract idea of the components of QuickCent.

Analogously to QuickEst [132], our QuickCent method relies on vectors of n *binary clues*. We associate to every node $v \in V$ a vector $\vec{x}_v = (x_v^1, x_v^2, \ldots, x_v^n) \in \{0,1\}^n$. The intuition is that the value of the $i-$th component (clue) x_v^i is an indicator of the presence ($x_v^i = 1$) or absence ($x_v^i = 0$) of an attribute signal of greater centrality for node v. Our method also considers the following $n+1$ sets of nodes:

$$\begin{aligned} S_1 &= \{v \in V \mid x_v^1 = 0\} \\ S_i &= \{v \in V \mid x_v^i = 0 \text{ and } x_v^{i-1} = 1\} \quad (2 \leq i \leq n) \\ S_{n+1} &= \{v \in V \mid x_v^n = 1\} \end{aligned}$$

That is, S_i corresponds to nodes that do not have the $i-$th attribute while having the previous one. For each one of the sets S_i, with $1 \leq i \leq n+1$, our method needs a quantity $\bar{f}_i$ which is a summary statistic of the centrality distribution of the nodes in set S_i. QuickCent must ensure that successive clues are associated with higher centrality values, thus we will have that

$$\bar{f}_1 < \bar{f}_2 < \cdots < \bar{f}_n < \bar{f}_{n+1}. \tag{3.1}$$

With the previous ingredients, the general estimation procedure corresponds to the following simple rule.

General QuickCent heuristic: *For node v, we iterate over the n clues, considering every value x_v^i. When we find the first i verifying that $x_v^i = 0$, we stop and output the value $\bar{f}_i$. If node v is such that $x_v^i = 1$ for every $i \in \{1, \ldots, n\}$, we output $\bar{f}_{n+1}$.*

Example 3.2.1 This is only a very simple example to exhibit the working of QuickCent, where we assume complete knowledge of the centrality values of all nodes. Let us consider the following network in Figure 3.1 of size 25 obtained as a random instance of linear preferential attachment, defined in Section 2.3.4 of the survey chapter. Table 3.1 displays only the non-zero values of in-degree and harmonic centrality in this network. A reasonable way to aggregate these values is to consider four sets $S_i, i = 1, 2, 3, 4$, with the following binary clues, $x_v^i = 1$ ($i = 1, 2, 3$) if and only if $\deg^{\text{in}}(v) > d_i$, with $d_1 = 0$, $d_2 = 3$ and $d_3 = 4$. With this choice, for simple centrality approximation it is natural to take, for example, the median of harmonic centrality on every set S_i as summary statistics, $\bar{f}_1 = 0$, $\bar{f}_2 = 1$, $\bar{f}_3 = 4.666$ and $\bar{f}_4 = 15.75$.

QuickCent provides a simple stopping rule: for each node, the search is finalized when the first clue with value 0 is found. Therefore, if our input is a network in which the vast majority of nodes is having similar and small centrality values –as it would be the case if the centrality were distributed according to a power law– the procedure is likely to stop the search early and give an estimate quickly. In this sense, the heuristic is *frugal*, given that in many cases it can output an estimate without passing over all the clues, or without using all the available information.

Up to this point, QuickCent remains similar to QuickEst. The reader can review the details of QuickEst in the book chapter by Hertwig et al (1999) [132]. The most critical

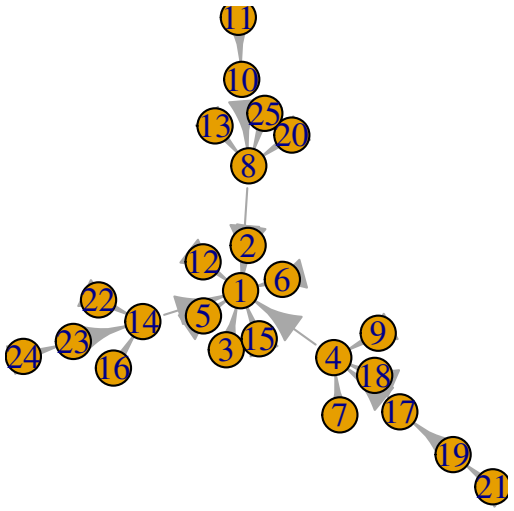


Figure 3.1: **A network randomly generated with linear preferential attachment.** Notice the formation of a central hub, and a vast majority of nodes with zero or low in-degree.

Node	1	4	8	10	14	17	19	23
In-degree	9	4	4	1	3	1	1	1
Harmonic	15.750	4.833	4.500	1.000	3.500	1.500	1.000	1.000
QC100	13.429	2.973	2.973	1.309	1.309	1.309	1.309	1.309
QC70	6.531	2.197	2.197	1.214	1.214	1.214	1.214	1.214

Table 3.1: **In-degree and harmonic centrality values for each node of the network from Figure 3.1.** Nodes that do not appear here have a zero in-degree and centrality. The last two rows correspond to QuickCent models described in Example 3.3.2. The number of decimal places is truncated to three with respect to the source.

aspects that distinguish QuickCent from QuickEst, as well as a specification of each part of the heuristic, are presented in the next section.

3.3 A QuickCent implementation

In this section, we propose an instantiation of our general QuickCent method, including a way to compute the clues x_v^i for every node v based on its in-degree in Section 3.3.1, and an efficient way to compute the summary statistic $\bar{f}_i$ of the centrality for every set S_i in Section 3.3.3. Section 3.3.2 makes explicit the assumptions on the structure of graphs that QuickCent requires to be a *ecologically rational heuristic* [132], i.e. the proper problem conditions that ensure a successful application of the heuristic, including that the centrality has a power-law distribution. Necessary concepts of the power-law distribution are introduced in Sections 2.2.3 and 2.2.5 from the survey chapter. Sections 2.3.3, and B.1 in the supplementary information annex of this chapter, detail the practical procedure we use to estimate the parameters of the power-law distribution from random samples.

3.3.1 Using the in-degree for the clues

Our approach for computing the binary clues is to employ a proxy variable related to the centrality by means of a monotonic function which ensures that Equation (3.1) holds. The idea is to use a proxy which should be far cheaper to compute than computing the actual centrality value. The proxy variable we chose is the in-degree of the node, that is, the number of neighbors of the node given by incoming arcs of the network. The intuition for this proxy is that greater in-degree will likely be associated with shorter distances, which likely increases the harmonic centrality. The in-degree is one of the most elementary properties of a node, and in many data structures it is stored as an attribute of the node (thus accessible in $O(1)$ time). The in-degree can itself be considered as a centrality measure [40]. For a node v we denote by $\deg^{\text{in}}(v)$ its in-degree.

Now, starting from a set of proportions $\{p_i\}_{i=1}^n$, where $0 \leq \cdots \leq p_i \leq p_{i+1} \leq \cdots \leq 1$, we can get the respective *quantile degree values* $\{d_i\}_{i=1}^n$. That is, if F is the cumulative distribution function (CDF) for the in-degree, then $d_i = F^{-1}(p_i)$ for each $i = 1, \ldots, n$. Then, we define the $i-$th clue for node v as

$$x_v^i = 1 \text{ if and only if } \deg^{\text{in}}(v) > d_i. \tag{3.2}$$

With this definition, the sets S_i are

$$\begin{array}{rcl} S_1 & = & \{v \in V \mid \deg^{\text{in}}(v) \leq d_1\}, \\ S_i & = & \{v \in V \mid d_{i-1} < \deg^{\text{in}}(v) \leq d_i\} \quad (2 \leq i \leq n), \\ S_{n+1} & = & \{v \in V \mid d_n < \deg^{\text{in}}(v)\}. \end{array}$$

Example 3.3.1 This type of clues was already used in Example 3.2.1. In fact, the quantile degree values $\{0, 3, 4\}$ used there can be obtained via the inverse of the in-degree CDF applied to the set of proportions $\{0.68, 0.84, 0.96\}$.

The final piece to apply QuickCent is to show how to compute the summary statistic $\bar{f}_i$ for every set S_i. We propose to compute $\bar{f}_i$ analytically as the median of each S_i based on estimating the parameters of a power-law distribution. This idea is developed in the next subsection.

3.3.2 Computing the summary statistic via a power-law distribution assumption

Our first assumption is the existence of a non-decreasing function g relating the in-degree and the centrality[1]. If there exists a function g satisfying this condition, then the quantiles in the centrality side are equivalent to the application of g on the same degree quantiles [138]. With this result, the quantile proportions can be specified according to characteristics of the centrality distribution, as it is explained in Section 3.3.3. In practice, and even more so considering that the in-degree is a discrete variable while the centrality is continuous, the object g is a relation rather than a function. More formally, let $\{C_i\}_{i=1}^n$ be the set of

[1]It is required an additional assumption –left continuity– which can be consulted at Hosseini (2010) [138].

quantile centrality values associated to the proportions $\{p_i\}_{i=1}^n$ that were used to compute the quantile degree values $\{d_i\}_{i=1}^n$ (see Equation (3.2)). Given the above assumption about g, we can rewrite the sets S_i as follows:

$$\begin{array}{rcl} S_1 & = & \{v \in V \mid g(\deg^{\text{in}}(v)) \leq C_1\} \\ S_i & = & \{v \in V \mid C_{i-1} < g(\deg^{\text{in}}(v)) \leq C_i\} \quad (2 \leq i \leq n) \\ S_{n+1} & = & \{v \in V \mid C_n < g(\deg^{\text{in}}(v))\} \end{array}$$

Our second assumption is that the centrality index that we want to estimate follows a power-law distribution. The introduction to this distribution is in Sections 2.2.3 and 2.2.5 from the survey chapter, while Sections 2.3.3, and B.1 in the supplementary information annex of this chapter, detail the practical procedure we use to estimate the parameters of the power-law distribution from random samples. We add this assumption motivated by the argument that QuickEst would have a *negative bias* [132], in the sense that it is a negative clue (or absent attribute) that stops this heuristic. Thus, a distribution such as the power law where most values are small (with mostly negative clues) and only a few high values exist (with mostly positive clues), would provide an optimal context for the performance of QuickEst, which is consistent with the finding that this heuristic predicts well the estimation behavior by some people on this kind of data [282]. Moreover, power-laws have a pervasive presence in many natural phenomena and magnitudes produced by humans too [206], although there has been some recent controversy on this topic [53, 19]. As we next show, our assumption of power-law distribution will allow us to use some particular properties to approximate the values $\{C_i\}_{i=1}^n$ used in the rewriting above, and then use them to efficiently compute the statistics $\{\bar{f}_i\}_{i=1}^{n+1}$ for every set S_i. In Section 3.4.3, we show some experiments to argue that these two assumptions of the heuristic are key to ensure its competitive accuracy.

3.3.3 Putting all the pieces together

Let $D = (V, A)$ be our input network, and recall that we are assuming that the centrality that we want to estimate for D follows a power-law distribution. Let $\hat{\alpha}$ be the estimate of the exponent parameter of the distribution (given by Equation (2.11) from Section 2.3.3 in the survey chapter), and $\hat{x}_{\min}$ be the estimate of the lower limit of the distribution (given by the minimization in Equation (2.12) from Section 2.3.3, estimated according to the method given in Annex B.1), which have been computed by considering a set of m nodes in V and their (real) centrality values. With all these pieces, we can compute the values $\{C_i\}_{i=1}^n$ associated to the proportions $\{p_i\}_{i=1}^n$ easily by using the equation

$$\int_{\hat{x}_{\min}}^{C_i} K x^{-\hat{\alpha}} dx = p_i$$

from which we get that

$$C_i = \hat{x}_{\min} \cdot (1 - p_i)^{\frac{1}{1-\hat{\alpha}}}.$$

Now, in order to compute the summary statistics $\{\bar{f}_i\}_{i=1}^{n+1}$, we will use the median of every set S_i. This median can be computed as follows. Given that we rewrote S_i as the set of

centrality values x such that $C_{i-1} \leq x \leq C_i$, then the median md_i of S_i must verify

$$\int_{md_i}^{C_i} Kx^{-\hat{\alpha}} dx = \frac{1}{2} \int_{C_{i-1}}^{C_i} Kx^{-\hat{\alpha}} dx$$

from which we obtain that

$$md_i = \left(\frac{(C_{i-1})^{1-\hat{\alpha}} + (C_i)^{1-\hat{\alpha}}}{2} \right)^{\frac{1}{1-\hat{\alpha}}} = \bar{f}_i \quad (2 \leq i \leq n) \tag{3.3}$$

Moreover, since the extreme points of the distribution are $x_{\min}$ (estimated as $\hat{x}_{\min}$) and ∞, the two remaining statistics $\bar{f}_1$ and $\bar{f}_{n+1}$ are computed as

$$\bar{f}_1 = \left(\frac{(C_1)^{1-\hat{\alpha}} + (\hat{x}_{\min})^{1-\hat{\alpha}}}{2} \right)^{\frac{1}{1-\hat{\alpha}}} \tag{3.4}$$

and

$$\bar{f}_{n+1} = 2^{\frac{1}{\hat{\alpha}-1}} \cdot C_n \tag{3.5}$$

We stress that with these formulas we compute the summary statistic $\bar{f}_i$ for each set S_i just by knowing the values $\{C_i\}_{i=1}^n$, which are computed by using only the values $\hat{\alpha}$, $\hat{x}_{\min}$, and the underlying vector of proportions $\{p_i\}_{i=1}^n$. This last element was chosen as the quantile probability values that produced equidistant points on the range of $\{\log(h(v)) \mid v \in V, h(v) \geq \hat{x}_{\min}\}$, that is, the set of vertices where the power-law is well defined for the harmonic centrality. Logarithmic binning is chosen to gauge the tail of the power-law distribution with higher frequency. The length n of the vector of proportions required to construct the clues (see Equation (3.2)) was chosen after pilot testing on each type of distribution. See Annex B.3, and Sections 3.4.3 and 3.4.4 for more details. The election of this vector is a way of adapting QuickCent to distinct centrality distributions. Research on possible improvements achievable by tuning this vector may be addressed in future work.

The last element we introduced in our procedure, is the use of an additional quantile centrality value $C_0 = \hat{x}_{\min}$, with the goal of spanning the centrality values $h(v) < \hat{x}_{\min}$ with greater accuracy. Since for this range of the vertex set the power-law distribution is no longer valid, the representative statistic $\bar{f}_0$ we have used is simply the empirical median of the harmonic centrality in the set of nodes v such that $\deg^{\text{in}}(v) \leq g^{-1}(\hat{x}_{\min})$. With this element, it turns out that, if we use a proportions vector $\{p_i\}_{i=1}^n$ of length n, the total number of medians $\{\bar{f}_i\}_{i=0}^{n+1}$ is $(n+2)$. In the code provided to produce the analyses of this paper [231], this element is optional (and activated by setting **rm**=True or **rms**=True). All the results in this paper were obtained with this centrality quantile and median.

Example 3.3.2 We continue revisiting Example 3.2.1. If we fix $x_{\min} = 1$, the exponent $\hat{\alpha}(1)$ that fits the complete distribution of centrality values, by using Equation (2.11) from Section 2.3.3 in the survey chapter, is 2.067. The set of proportions shown in Example 3.3.1 comes from evaluating the centrality CDF on the set of points $\{1, 2.506, 6.283\}$, which correspond to $x_{\min}$ and two points ($n = 2$) that in logarithmic scale turn out to be equidistant to the minimum and maximum of the set $\{\log(h(v)) \mid v \in V, h(v) \geq \hat{x}_{\min}\}$, the (log) centrality domain of the given network where the power law is valid. From these parameters and the expressions shown in this section, one can get the medians required by QuickCent to

make estimates. These can be examined in Table 3.1, corresponding to the model **QC100**, which has a MAE (mean absolute error) over the whole digraph of $3.606e-01$. A more interesting case may be computed when $\hat{\alpha}(1)$ is derived from a random sample of the centrality distribution. For example, by taking a sample without replacement of size 70% one may get an exponent estimate of $\hat{\alpha}(1) = 2.477$, which has a MAE of $6.948e-01$ and QuickCent estimates that can be examined in the model **QC70** in Table 3.1.

This completes all the ingredients for our instantiation of QuickCent, as we have the values for the clues $(x_v^1, x_v^2, \ldots, x_v^n)$ computed from the in-degree of the node v, plus the values $\{d_i\}_{i=1}^n$ as shown in Equation (3.2), and also the summary statistics $\{\bar{f}_i\}_{i=0}^{n+1}$ for each set S_i, which are the two pieces needed to apply the heuristic.

3.4 Results

In the present section, we show the results of applying *QuickCent* either on synthetic data or on some empirical networks, and we compare it with alternative procedures for centrality estimation. We first show the comparison of QuickCent with other methods when applied on synthetic networks, considering accuracy and time measurements in Sections 3.4.1 and 3.4.2. The synthetic network model corresponds to the preferential attachment (PA) growth model introduced in Section 2.3.4 from the survey chapter. Section 3.4.3 reviews the output of QuickCent on null network models where its accuracy is not as good relative to other methods, with the aim of showing that the two assumptions of QuickCent (Section 3.3.2) are jointly required as a necessary condition for the competitive performance of this heuristics. The same benchmark presented for the synthetic case was applied to the empirical datasets, and the results are shown in Section 3.4.4. The experiments to check the fulfillment of QuickCent assumptions by the different networks are shown in Sections B.2, B.4, B.5 and B.6 from the supplementary information annex. In all our experiments we consider harmonic centrality as the target to estimate. The number of nodes chosen for the synthetic networks experiments is $10,000$ and 1000 for the null models, with the aim of accelerating the bootstrap computations to check the assumptions of QuickCent on each network. Similar sizes were searched for when choosing the tested empirical networks. These are not really big numbers compared with modern networks. We select these numbers as we need to be able to compute the exact value of the harmonic centrality for all nodes in the graph, in order to compare our estimations with the real value, and regard it to be enough for a first assessment of the heuristic.

Experiments specifications

The norm that we employed to summarize the error committed on each node is the mean absolute error (MAE). This measure is preferable to other error norms, such as the Root mean squared error, because the units of the MAE are the same as the quantity under consideration, in this case, the harmonic centrality. On the other hand, since the MAE can be understood as the *Minkowski* loss with $\mathcal{L}_1$ norm for the regression of the variable of interest, and in this case, it is known that the solution is given by the conditional median [35], it is reasonable to use the MAE when the summary statistic chosen is the median of each centrality interval.

Finally, all the experiments were performed on the R language [237] with igraph library [76] for graph manipulation, and ggplot2 library [288] to produce the plots.

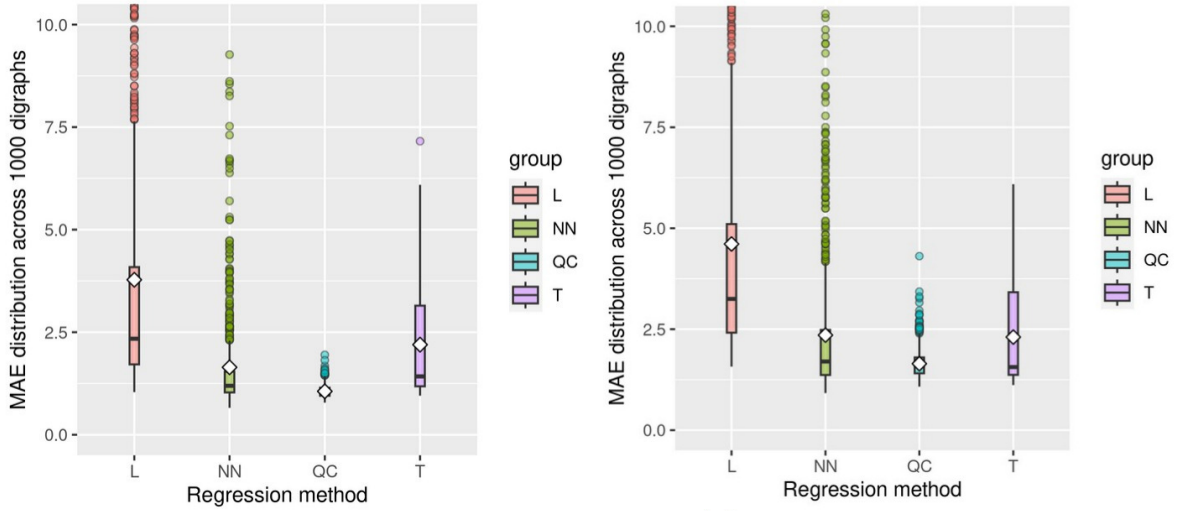
3.4.1 Comparison with other methods

In this section, we compare the performance of known existing regression methods with QuickCent. This exercise allows us to evaluate the potential uses and applications of our proposal. Specifically, whether it can deliver reasonable estimates, in relation to alternative solutions for the same task. This is not a trivial matter, considering that QuickCent is designed to do little computational work of parameter estimation and output production, possibly with limited training data, while common alternative machine learning (ML) methods generally perform more complex computations. For a fair comparison, all other methods use only the in-degree as an explanatory variable. In rigor, QuickCent is able to produce the estimates only from the binary clues, without using the in-degree.

The competing methods considered are linear regression (denoted by L in plots), a regression tree (T) [236, 291] and a neural network (NN) [250], which are representatives of some of the most known machine learning algorithms. We used *Weka* [291] and the *RWeka* R interface [137] to implement T and NN using default parameters. In the literature, there is previous work specifically tailored to centrality estimation using ML methods, but for other centrality indices beyond harmonic centrality. In particular Brandes and Pich study specific estimations for *closeness* and *betweenness* centrality [47]. It would be interesting to compare our method with the one proposed by Brandes and Pich [47], but this would amount to changing and adapting their method to harmonic centrality. We leave this adaptation and further comparison as future work.

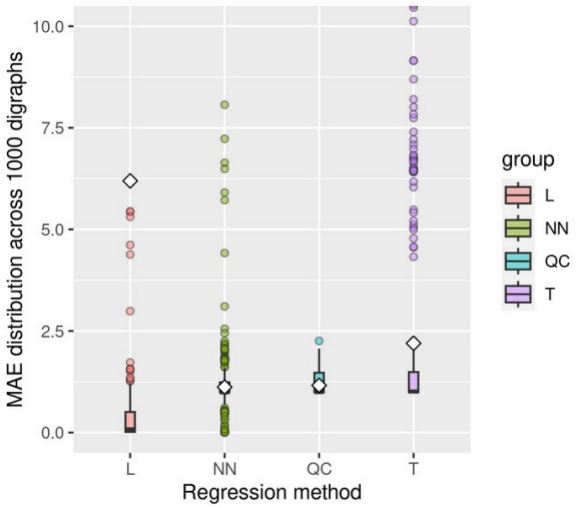
The results of this experiment are shown in Figure 3.2 with a training size of 10 % and Figure 3.3 with a training size of 100 %, where the test set is always the entire digraph. The two training sizes are studied with the goal of assessing the impact of scarce data on the distinct estimation methods, by contrasting a full versus a scarce data scenario. In the figures, it can be seen that QuickCent (QC) produces the lowest MAE errors of all the methods, either in terms of the IQR length, or the mean and outliers, for PA exponents 1 and 0.5. As noted in the Annex B.2, these are the cases where the centrality distribution has a better fit by a power-law model, but it is anyway a remarkable result considering the error committed by fixing $x_{\min} = 1$, see Annex B.3. Examining the MAE medians of these simulations in Table 3.2, one sees that the QC median is at the level of the most competitive ML methods in the simulations, NN and T, sometimes being the best of the three depending on the experiment. However, for exponents 1 and 0.5, where the power-law is present, the good thing about QC is that the upper quantiles and even the outliers remain low compared to other methods (see Figure 3.2 and Figure 3.3).

Thus, the main takeaway is that QC, when its assumptions are fulfilled, is able to produce estimates at the same level as much more complex ML methods, with likely lower variance. This fact is consistent with the argument given by Brighton and Gigerenzer [49] claiming that the benefits of simple heuristics are largely due to their low variance. The argument relies on the decomposition of the (mean squared) error into *bias*, the difference between the



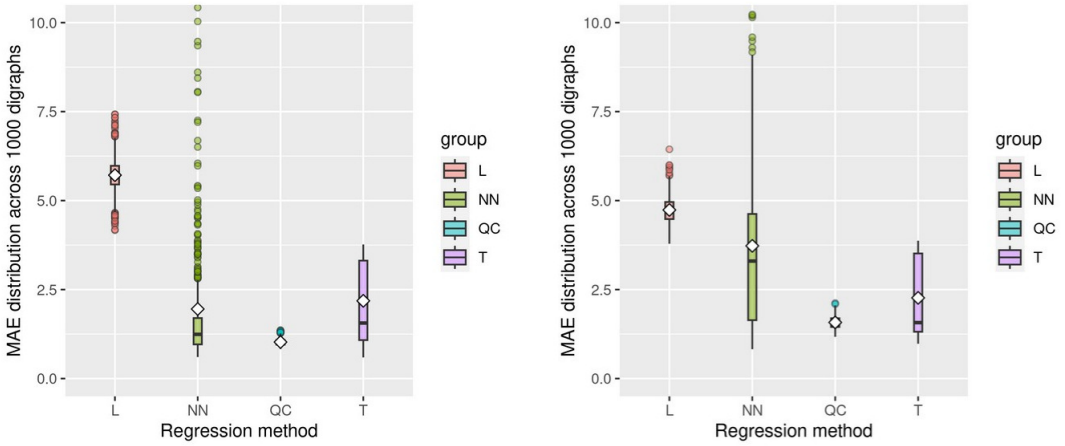
(a) PA exponent= 1

(b) PA exponent= 0.5



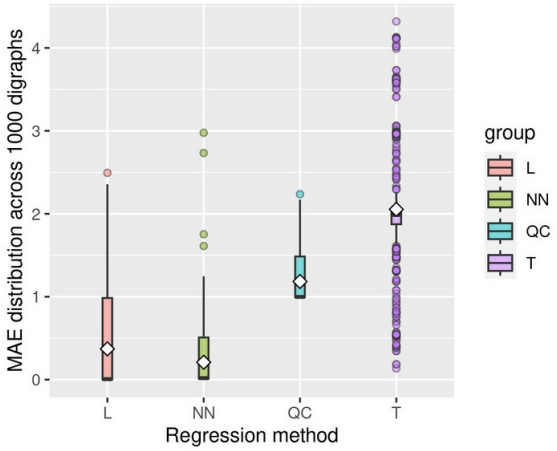
(c) PA exponent= 1.5

Figure 3.2: **Benchmark with other ML methods for different exponents of PA digraph instances and 10 % of training size.** For each regression method, there is a boxplot showing the MAE distribution. Each boxplot goes from the 25−th percentile to the 75−th percentile, with a length known as the *inter-quartile range* (IQR). The line inside the box indicates the median, and the rhombus indicates the mean. The whiskers start from the edge of the box and cover until the furthest point within 1.5 times the IQR. Any data point beyond the whisker ends is considered an outlier, and it is drawn as a dot. For display purposes, the vertical limit of the plots has been set to 10, since the highest MAE outliers of NN or L, depending on the PA exponent, blur the details of the model performance. Observe that QuickCent is in general reliable (has low variance), and has the best accuracy for the PA exponents (1 and 0.5) where the harmonic centrality is well approximated by a power-law.



(a) PA exponent= 1

(b) PA exponent= 0.5



(c) PA exponent= 1.5

Figure 3.3: **Benchmark with other ML methods for different exponents of PA digraph instances and 100 % of training size.** For each regression method there is a boxplot showing the MAE distribution. Each boxplot goes from the 25−th percentile to the 75−th percentile, with a length known as the *inter-quartile range* (IQR). The line inside the box indicates the median, and the rhombus indicates the mean. The whiskers start from the edge of the box and cover until the furthest point within 1.5 times the IQR. Any data point beyond the whisker ends is considered an outlier, and it is drawn as a dot. For display reasons, the vertical limit of the two first plots was set to 10, since the highest MAE outliers of NN or L, depending on the PA exponent, blur the details of the model performance. Observe that QuickCent is in general reliable (has low variance), and has the best accuracy for the PA exponents (1 and 0.5) where the harmonic centrality is well approximated by a power-law.

PA β	L10	NN10	QC10	T10	L100	NN100	QC100	T100
1	2.341	1.194	1.040	1.422	5.711	1.242	1.009	1.560
0.5	3.249	1.699	1.576	1.561	4.704	3.300	1.578	1.571
1.5	0.079	0.991	0.996	1.009	0.006	0.018	0.997	1.986

Table 3.2: **Medians of the MAE distribution across 1000 digraphs.** These estimates are computed from the same simulationes displayed in Figure 3.2 and Figure 3.3. The suffix of each method abbreviation corresponds to the size of the training size used. The exponent β corresponds to the exponent of the preferential attachment growth (Section 2.3.4 from the survey chapter). The number of decimal places is truncated to three with respect to the source.

average prediction over all data sets and the desired regression function, and *variance*, the extent to which the estimates for individual datasets vary around their average [35]. Thus, along the range of the bias-variance trade-off of models, simple heuristics are relatively rigid models with high bias and low variance, avoiding the potential overfitting of more complex models.

By examining the contrast of the outliers between Figure 3.2 and Figure 3.3, it can be noticed that QC suffers the least impact from scarce data. In the case of L and NN, they show a similar pattern for power-law centralities (PA exponents 1 and 0.5). They have medians that are lower for the 10 % training size than those obtained with the whole network. Since there are only a few large values in the entire graph, when the training sample gets smaller, the sample values have a better linear fit, in comparison to larger samples. Therefore, a linear model adjusted to some small sample provides a good fit to the small-to-moderate centrality size nodes, which is the case for most of the nodes. This also explains the presence of higher outliers in the 10 % training size. On the other hand, the behavior of the regression tree is more similar to that of QuickCent.

3.4.2 Time measurements

Elapsed time was the other distinct aspect of the method's performance. The time cost is a critical feature of any approximation method because it measures the tradeoff between accuracy and cost. Elapsed time measurements were taken in the experiments shown in Section 3.4.1, and the results are displayed in Table 3.3. These times consider the training and inference time for each method, without including any centrality computation. In the case of QuickCent, the computation of the proportions vector is not considered for the elapsed times. The reason is that this computation is not directly part of the heuristics, and there is also the possibility of using a prototypical vector for a given type of centrality distribution.

From the table, we can see that the elapsed time of QC is in the middle range of the compared methods. The linear regression has the lowest times, around one order of magnitude faster than QuickCent and the regression tree, and the neural network has the highest elapsed time. Note also that there is no significant time difference between the 10%-case and the 100%-case for QC. This can be explained by the fact that the differences in sample sizes only

	L10	QC10	T10	NN10	L100	QC100	T100	NN100
Mean	2	76	23	98	3.9	77.5	28.0	760.9
St. Dev.	0.25	2.95	2.98	2.16	0.35	3.27	4.66	9.88

Table 3.3: **Mean and standard deviation of elapsed time in milliseconds over** 1000 **digraphs with PA exponent 1.** The suffix of each method abbreviation corresponds to the size of the training set used.

affect the number of terms in the sum in Equation (2.11) (Section 2.3.3 in the survey chapter) when estimating the exponent $\hat{\alpha}$, and summing a list of values is an extremely simple and quick procedure.

Based on these results, we conjecture that QuickCent has the lowest time complexity among the tested methods. Among the computations that QuickCent performs, the most expensive ones correspond to the selection problem of finding the median of the lowest centrality values (Section 3.3.3), plus the quantile degree values (Section 3.3.1). The procedure used to compute the proportions vector (Section 3.3.3), which is not considered in the elapsed time measurements, also relies on solving the selection problem (of the maximum of the set of centrality values) and sorting of the centrality values set (to find the proportions). These problems may be solved in linear time [75] on the input size, that is, linear on the network size $\mathcal{O}(\mid V \mid)$. In contrast to the highly optimized R implementations for L, T, and NN, we considered only a naive implementation of QuickCent without, for example, architectural considerations. With these improvements such as using more appropriate data structures, these times could still be improved. We left as future work the construction of an optimized implementation for QuickCent.

3.4.3 Networks defying QuickCent assumptions

Up to this point, we have mainly seen examples of networks where QuickCent exhibits quite good performance compared to competing regression methods. In order to give a full account of QuickCent capabilities and its *ecological rationality* [132], one should also have an idea of the networks where its accuracy deteriorates. To accomplish this, we will look at the two assumptions of QuickCent, namely, the power-law distribution of the centrality, and its monotonic map with the in-degree, to show that they are jointly required as a necessary condition for the competitive performance of the heuristic. Our approach here is to work with two null network models, each acting as a negation of the conjunction of the two assumptions, which provide strong evidence for this claim.

Response to the loss of the monotonic map

Our first null model is a scale-free network built by preferential attachment, just as in the previous experiments, but after a *degree-preserving randomization* [194] of the initial network, which is simply a random reshuffling of arcs that keeps the in- and out-degree of each node constant. The aim is, on the one hand, to break the structure of degree correlations found

in preferential attachment networks [185, 297], which may be a factor favoring a monotonic relationship between in-degree and harmonic centrality, and on the other hand, to maintain a power-law distribution for the harmonic centrality by preserving the degree sequence of nodes in the network. This last feature does not ensure that the centrality distribution is a power-law, since randomization also affects this property. Annex B.4 displays the results of the experiments performed to check the assumptions of QuickCent on the randomized networks, showing that these networks satisfy them. Finally, Figure 3.4 shows the impact of randomization on each regression method. This is an experiment where 1000 PA networks (exponent 1) of size 1000 were created, and the four ML methods used in Section 3.4.1 were trained on each network with samples of size 30 % of the total node set, using only the in-degree as the predictor variable for the harmonic centrality. The same procedure was run on each network after applying degree-preserving randomization on 10000 pairs of arcs. The plot shows that the randomization has a similar impact on the performance loss of each method, which is an expected result due to the fact that the only source of information used by each method, the in-degree, becomes less reliable due to the weaker association with harmonic centrality thanks to the arc randomization. Since QuickCent was the most accurate of the methods tested on the initial PA networks, it appears to be also one of the methods most affected methods by the randomization.

Response to the loss of the power-law distribution of centrality

Our second null model is the directed Erdös-Rényi graph model [92, 41, 157], and is chosen with the aim of gauging the impact of losing the power-law distribution of the centrality while maintaining the monotonic map from in-degree to centrality. This model is known to have a Poisson degree distribution [41], a behavior very different from a heavy-tailed distribution, and according to our simulations (see Annex B.5), it turns out to be ideal for our purposes. We choose connection probabilities that ensure a unimodal distribution for centrality and a strong correlation with in-degree, i.e. with a mean in-degree greater than 1 [157]. In order to get a fair control on the performance of QuickCent, we have taken two empirical digraphs that satisfy the given condition of the mean in-degree, with node sets of size near 1000, just to accelerate the bootstrap p-value computations. The networks are extracted from the KONECT database [176][2], and their meta-data is shown in Table 3.4. The fields N and $\overline{\deg}^{\text{in}}$ given in this table, are used to determine the network size and the connection probability used to instantiate the respective ER digraphs from the identity $\overline{\deg}^{\text{in}} = p \cdot (N - 1)$.

Finally, in Figure 3.5 we can see the results of an experiment analogous to the one with the first null model, that is, there are 1000 iterations where the same four ML methods were trained on each network, two ER graphs with the two connection probabilities and sizes given by the two empirical/control networks, with random samples of size 30 % of the total node set, using only the in-degree. Since the unimodal distribution of ER digraphs is very different from a power-law, in this experiment we have taken the approach of using the parameter $\hat{x}_{\min}$ estimated by the method reviewed in Annex B.1, as well as the search space restriction explained there, instead of a fixed lower limit as in the previous experiments. Now, by comparing the two plots in Figure 3.5, one can observe a noticeable difference in

[2] http://konect.cc/

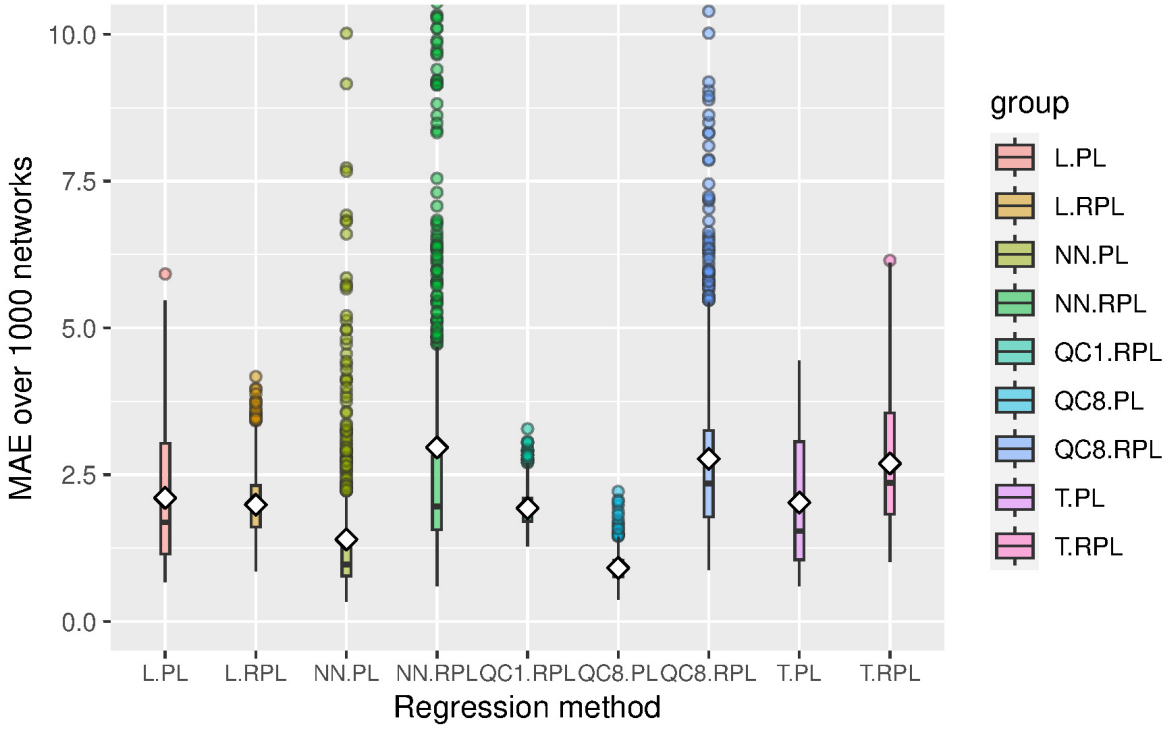


Figure 3.4: **Effect of randomization on different ML methods using 30 % of the training size.** Each boxplot group is labeled with the name of the ML method, a dot, and the type of network on which the estimates are made ('PL' for the initial PA network, 'RPL' for the network after randomization). QC8 corresponds to QuickCent with a proportion vector of length 8, and analogously for QC1. For each regression method, there is a boxplot representing the MAE distribution. Each boxplot goes from the 25−th percentile to the 75−th percentile, with a length known as the *inter-quartile range* (IQR). The line inside the box indicates the median, and the rhombus indicates the mean. The whiskers start from the edge of the box and extend to the furthest point within 1.5 times the IQR. Any data point beyond the whisker ends is considered an outlier, and it is drawn as a dot. For display reasons, the vertical limit of the plots was set at 10, since the highest MAE outliers of NN, make blur the details of the model performance. The arc randomization impacts the monotonic map from in-degree to harmonic centrality, having a detrimental performance effect on almost every regression method.

Name	N	$\bar{\deg}^{\text{in}}$	Corr	Arc meaning	Ref.
moreno_blogs	990	19.21	0.872	Blog hyperlink	[2]
subelj_jung-j	2208	62.81	0.808	Software dependency	[267]

Table 3.4: **General description of the two empirical control networks.** The fields in the table are the dataset name, the number of nodes with positive in-degree (N), the mean in-degree of nodes with positive in-degree ($\bar{\deg}^{\text{in}}$), the Spearman correlation between the positive values of in-degree and harmonic centrality (Corr), the meaning of the arcs, and the original reference. The name corresponds to the *Internal name*field in the KONECT database. To access the site to download the dataset, append the internal name to the link *http://konect.cc/networks/*.

the behavior of QuickCent in the two cases. While QuickCent achieves an average accuracy relative to other regression methods on the control networks with centrality distributions that are more or less close to heavy-tailed, on ER digraphs with similar characteristics to the controls, QuickCent consistently performs worse than other methods. The performance of QuickCent in these plots corresponds to the best possible for each network as a function of the length of the proportions vector, denoted by the number after 'QC'. This output is also consistent with the difference in p-values of the power-law fit between the control networks and 1000 instances of the ER models, reported in Table B.5 from the supplementary information annex. These results reveal the critical importance of the centrality distribution of the data set for the proper functioning of QuickCent.

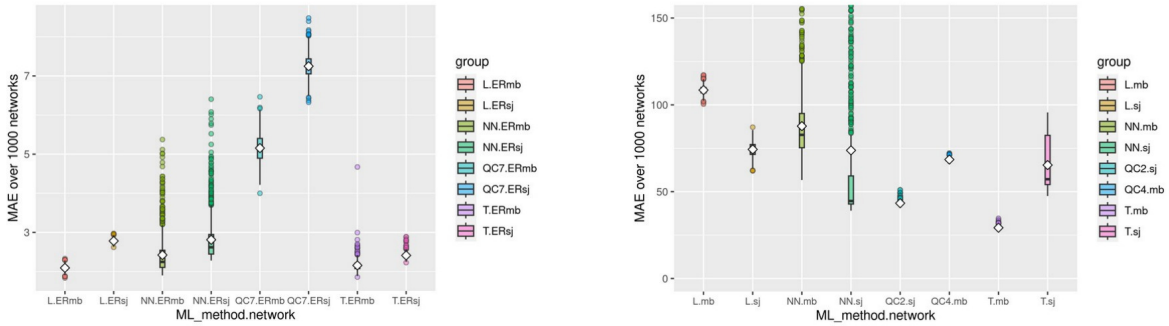
On the other hand, all of the methods exhibit better performance on the ER digraphs than on the corresponding control network, probably due to less heterogeneity in the values to be predicted on the former. Finally, as a side note for working on empirical network datasets, for general networks it should be more accurate to use the fitted value of $\hat{x}_{\min}$ than a fixed value, although this depends on the variability range existing on the values less than $\hat{x}_{\min}$, which may introduce potentially large contributions to the estimation error. Observe that there is an additional computational overhead due to the calculation of $\hat{x}_{\min}$.

3.4.4 Experiments with empirical networks

In this section, we present the performance of QuickCent on some real network datasets of similar size to the synthetic networks already tested, also in comparison to other machine learning methods. These results are only a first glimpse of the challenges this heuristics may encounter when dealing with real datasets, and they should also be considered as a proof of concept.

We selected five datasets, all of them extracted from the KONECT network database [176][3], a public online database of more than a one thousand network datasets. The criteria for selecting the networks were that, besides being similar in size to our synthetic networks (10000 nodes), each network had a distinct meaning, i.e., networks representing distinct

[3]http://konect.cc/



(a) ER networks (b) moreno_blogs and subelj_jung-j

Figure 3.5: **Effect of centrality distribution on different ML methods using 30 % of training size.** Each boxplot group is labeled with the name of the ML method, a dot, and the type of network on which the estimates are made ('mb' for moreno_blogs, 'sj' for subelj_jung-j, 'ERmb' for the ER digraph created with the parameters of moreno_blogs, and analogously for 'ERsj'). The number after 'QC' is the length of the vector of proportions used by that method, corresponding to the best accuracy for the respective network. For each regression method, there is a boxplot representing the MAE distribution. Each boxplot goes from the 25−th percentile to the 75−th percentile, with a length known as the *inter-quartile range* (IQR). The line inside the box indicates the median, and the rhombus indicates the mean. The whiskers start from the edge of the box and extend to the furthest point within 1.5 times the IQR. Any data point beyond the whisker ends is considered an outlier, and it is drawn as a dot. For display reasons, the vertical limit of the control network plot has been set at 150, as the highest MAE outliers of NN blur the details of the model performance. The comparison between the two plots, reveals the critical importance of the centrality distribution for QuickCent performance. In the left plot QuickCent presents a relative bad performance, while at the right it has a competitive average accuracy.

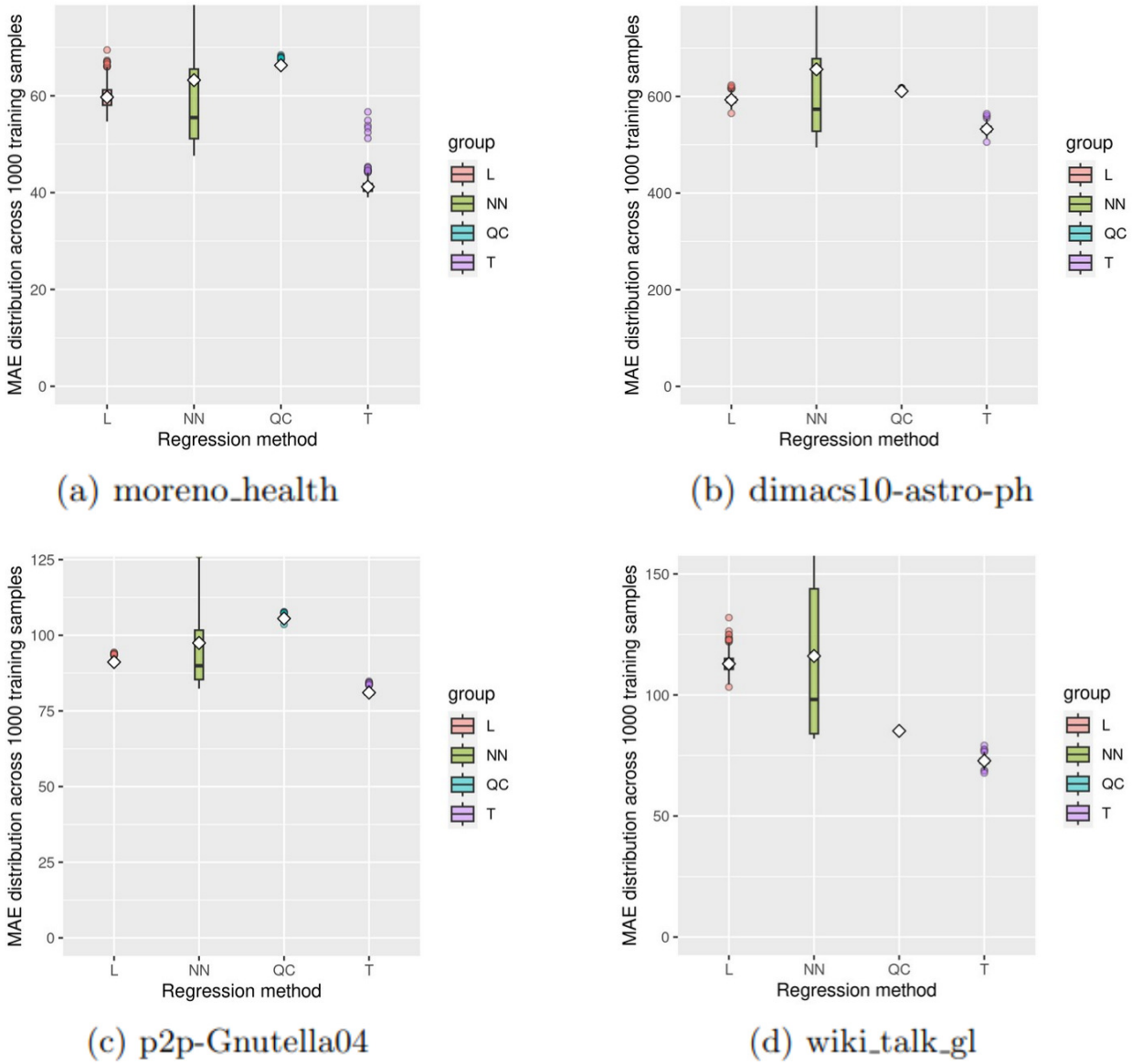


Figure 3.6: **Performance of QuickCent against known ML algorithms on each empirical dataset.** The competing algorithms are the same as in Section 3.4.1, that is, a linear regression (L), a neural network (NN), and a regression tree (T), all with default parameters. Each point from each boxplot is the MAE of the respective model trained with a random sample of nodes of size 10 % of the total, and all the samples come from the same respective network. The white rhombus in each boxplot is the mean of the distribution. Notice that QuickCent has, in general, an acceptable accuracy compared to other methods, although not as good as that obtained for synthetic datasets (see, for example, Figure 3.2).

Name	Dir.	N	m	Edge meaning	Reference
moreno_health	Y	2539	12969	Friendship	[205]
dimacs10-astro-ph	N	16046	121251	Co-authorship	[209]
dblp-cite	Y	12590	49759	Citation	[183]
p2p-Gnutella04	Y	10876	39994	Host Connection	[246]
wiki_talk_gl	Y	8097	63809	Message	[268]

Table 3.5: **General description of the five empirical network datasets.** The fields in the table are the dataset name, whether the network is directed, the number of nodes (N), the number of edges (m), the meaning of the edges, and the original reference . The name corresponds to the *Internal name* field in the KONECT database. To access the site to download the dataset, append the internal name to the link *http://konect.cc/networks/*.

systems from different contexts. General descriptors of these datasets are displayed on Table 3.5. There, we can see that we have selected: a social network of friendships among students created from a survey [205], a co-authorship network from the *astrophysics* section of arXiv from 1995 to 1999 [209], a citation network of publications from DBLP [183], a network of connected Gnutella hosts from 2002 [246], and the communication network of messages among users from the Galician Wikipedia [268]. See Annex B.6 to review the experiments of assumptions verification on these empirical datasets.

We end this section with several plots in Figure 3.6 showing the performance of QuickCent compared to the same ML algorithms from Section 3.4.1, all of them trained with samples of size equal to the 10 % of each dataset. The feature of QuickCent having the smallest error dispersion observed in the synthetic datasets is also observed in this case. The QC performance, although not as good as in the synthetic datasets, is competitive with the other ML methods, and even better than an important number of instances of the neural network for the dimacs10-astro-ph and wiki_talk_gl datasets. These results are obtained with a length of the proportions vector equal to 2, which delivers the best performance found among several vector lengths tested, in contrast to the larger length of 8 used in the synthetic case. These two differences with the synthetic case, support the hypothesis that the overall goodness of the power-law fit found by QC is better for the synthetic distributions than for the empirical ones. Finally, it is noteworthy that in these two datasets either QC or T, and in general the regression tree for all the datasets, obtain the best accuracy beating more flexible methods such as NN, considering that these methods provide a limited number of distinct output values.

3.5 Discussion and future work

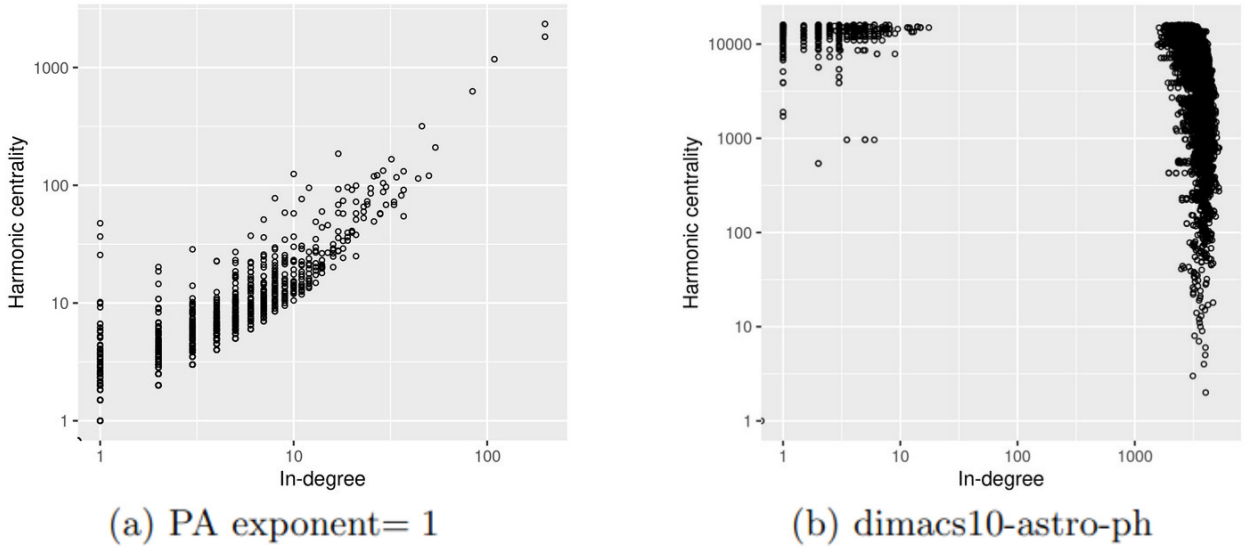
In this section, we analyze the results presented in the last section. We start with a summary of the results, and then the discussion is mainly centered on the type of network patterns on which the performance of QuickCent is based. We end up with a series of ideas for future work and concluding remarks.

Summary of results. The results presented in Section 3.4 show that QuickCent can be a competent alternative to perform a regression on a power-law centrality variable. The method generates accurate and low variance estimates even when trained on a small -10%- proportion of the dataset, comparable in precision to some more advanced machine learning algorithms. Its accuracy is available at a time cost that is significantly better than one of the machine learning methods tested, namely, the neural network. In this sense, QuickCent is an example of a simple heuristic based on exploiting regularities present in the data, which can be a competitive alternative to more computationally intensive methods.

The patterns on which QuickCent relies. An interesting question is why our initial attempt to approximate an expensive centrality sensitive to size and density by a cheap density measure is successful, at least for the network cases tested. The same question framed in terms of the QuickCent method, would be why the two method assumptions, the power-law of harmonic centrality, and the strong correlation between harmonic centrality and in-degree, do hold for power-law, or more specifically for some preferential attachment networks. It was already mentioned that while in-degree and PageRank centrality of a digraph obey a power-law with the same exponent [188], we do not know results describing the distribution of harmonic centrality on digraphs. A possible intuition for the scale-free behavior of harmonic centrality observed on PA networks (Figure B.1 in the supplementary information annex), may come from the motivation for the harmonic centrality given by Marchiori and Latora [193, 178]. The reciprocal shortest-path distances are used to informally define the efficiency of communication between nodes in a network. Therefore, it is reasonable to think that the scale-free degree distribution induces an analogous distribution in the efficiency to receive information sent.

The correlation and the monotonic relationship between harmonic centrality and in-degree, is strongly favored by the network generation mechanism. There is converging evidence showing that preferential attachment, which in its usual formulation requires global information about the current degree distribution, can be the outcome of link-creation processes guided by the local network structure, such as a random walk adding new links to neighbors of connected nodes, or in simple words, meeting friends of friends [143, 277]. The reason is that the mechanism of choosing a neighbor of a connected node makes those higher-degree nodes more likely to be chosen by the random walk, which in turn makes more paths lead to them. That is, the local density could indeed reflect the access to larger parts of the network. See also Section 2.3.6 from the survey chapter. Of course, preferential attachment is not the only mechanism capable of producing scale-free networks [172, 299, 86], and the distinct generative mechanisms may engender or not, a stronger relationship between density and size in the resulting network. This insight may be the reason why the monotonic relationship between harmonic centrality and in-degree is more apparent in the preferential attachment model than in some of the empirical networks, as Figure 3.7 shows.

The reasons explaining why and when fast and frugal heuristics work is an active research problem [49, 134]. This question is not addressed in this paper, but some related results are mentioned next. It has been claimed and tested by simulations that the QuickEst estimation mechanism works well on power-law distributed data, but poorly on uniform data [282]. In the case of QuickCent, this dependence on the power-law distribution is reinforced by



(a) PA exponent= 1 (b) dimacs10-astro-ph

Figure 3.7: **Scatterplot of in-degree versus harmonic centrality for a synthetic and an empirical network.** The plot on the left is obtained with a PA exponent of 1, and the plot on the right is that of the dimacs10-astro-ph dataset. The axes are in logarithmic (base 10) scale. Observe that the monotonic relation between in-degree and harmonic centrality, is much more clear in the synthetic network (left) that in the empirical dataset (right), which impacts the performance of QuickCent (see Figure 3.4).

the parameter construction of this method (see Section 3.3). Another fact given by the definition of our clues that has been pointed out as a factor favoring simple strategies is the correlated information [179], that is, information found early in the search is predictive of information found later. In the case of QuickCent, this simply means that since all clues are based on the in-degree, and this number can only belong to one of several disjoint real intervals, information from additional clues will not provide contradictory evidence once the heuristic has terminated its search. Other tasks where it is necessary to weigh the contribution of many possibly contradicting variables may present a more challenging context for these simple heuristics.

Future work. The insight described above about the monotonic map between the in-degree and the harmonic centrality on networks generated by preferential attachment nurtures the conjecture that QuickCent may be better suited to, for example, networks with an information component such as the Internet or citations, which can be well approximated by this growth mechanism [20, 148, 277], than to more pure social networks such as friendships [143, 53]. There is evidence that, if one assumes that some nodes to form links are found uniformly at random, while others are found by searching locally through the current structure of the network, it turns out that the more pure social networks appear to be governed largely through random meetings, while others like the World Wide Web and citation networks involve much more network-based link formation [143]. See also Sections 2.3.6 and 2.3.4 from the survey chapter. Testing this hypothesis on a large corpus of diverse empirical network datasets is an interesting question for future work.

Applying QuickCent to other types of networks or centrality measures is not a direct task, since, depending on the type of network considered, degree and centrality may be strongly or weakly related. We plan to address these extensions in future work, where one possible line of research is to formulate the problem of finding the proportion quantiles as that of obtaining an optimal quantizer [118]. There is some resemblance between our problem of finding the quantiles minimizing the error with respect to some distribution and that of finding the optimal thresholds of a piecewise constant function minimizing the distortion error of reproducing a continuous signal by a discrete set of points. On the other hand, QuickCent requires an explanatory variable that is correlated with the fitted variable to construct the clues. Future work should deal with extensions and flexibility of the clues employed, trying other clues or new ways to integrate different clues. The idea raised in our work of using a local density measure to approximate expensive size-based centrality indices could be generalized in order to be valid on more general networks, for example, by using a more general notion of local density than the in-degree, such as a modified degree measure to ensure minimum overlapping between spreading regions [175], or spreading indices based on the degree and neighbors' degree [29].

3.6 Conclusion

The results of this paper are a proof of concept to illustrate the potential of using methods based on very simple heuristics to estimate some network centrality measures. Our results show that QuickCent is comparable in accuracy to the best-competing methods tested, with the lowest error variance, even when trained on a small proportion of the dataset, and all this at intermediate time cost relative to the other methods using a naive implementation. We give some insight into how QuickCent exploits the fact that in some networks, such as those generated by preferential attachment, local density measures, such as the in-degree, can be a good proxy for the size of the network region to which a node has access, opening up the possibility of approximating centrality indices based on size such as the harmonic centrality.

Chapter 4

Modularity of food-sharing networks minimises the risk for individual and group starvation in hunter-gatherer societies

Abstract

It has been argued that hunter-gatherers' food-sharing may have provided the basis for a whole range of social interactions, and hence its study may provide important insight into the evolutionary origin of human sociality. Motivated by this observation, we propose a simple network optimization model inspired by a food-sharing dynamic that can recover some empirical patterns found in social networks. We focus on two of the main food-sharing drivers discussed by the anthropological literature: the reduction of individual starvation risk and the care for the group welfare or egalitarian access to food shares, and show that networks optimizing both criteria may exhibit a community structure of highly-cohesive groups around special agents that we call hunters, those who inject food into the system. These communities appear under conditions of uncertainty and scarcity in the food supply, which suggests their adaptive value in this context. We have additionally obtained that optimal welfare networks resemble social networks found in lab experiments that promote more egalitarian income distribution, and also distinct distributions of reciprocity among hunters and non-hunters, which may be consistent with some empirical reports on how sharing is distributed in waves, first among hunters, and then hunters with their families. These model results are consistent with the view that social networks functionally adaptive for optimal resource use, may have created the environment in which prosocial behaviors evolved. Finally, our model also relies on an original formulation of starvation risk, and it may contribute to a formal framework to proceed in this discussion regarding the principles guiding food-sharing networks.

4.1 Introduction

Food-sharing and its relevance for the evolution of human cooperation. Human beings are social animals that have built many forms to organize their life in community due to their remarkable ability to coordinate and jointly work to attain common goals. A fingerprint of these behaviors can be tracked in the myriad types of social networks human beings develop, such as family, friends, and communities of interest. Namely, it has been argued that a distinctive attribute of social networks is that they have communities or modules, that is, cohesive groups of agents more densely connected among them than with the rest of the network. This feature may help to explain other special attributes of social networks [212].

One of the key factors for the successful adaptation of the human species to almost every habitat on the planet along its evolutionary history [162], is the unique complexity of its patterns of food sharing, which extends well beyond infancy lactation to the whole life and

across adults and families [124]. These patterns may have emerged as a response to a greater offspring dependency associated with the development of a larger brain requiring a nutrient-rich diet, more difficult feeding strategies, and foraging skills that may not develop until late in life, in contrast to other wild primates having more predictable diets [145, 124, 154]. The large human brains may have had a central role both to the formation of modular social networks enabling adaptive collective computation, to the optimization of foraging behaviors [129]. These behaviors ensuring the most elemental subsistence may have given the material basis for inter-generational accumulation of cultural innovations [201], and the whole range of social interactions carried out by hunter-gatherer bands, such as cooperative breeding [170, 133], establishing political alliances and reinforcing social bonds [222, 163, 136, 241], labor sharing [245, 15] and costly signaling or show-off for mating purposes [258, 131]. Therefore, investigating the evolutionary formation of food-sharing networks may yield important insights for the study of social dynamics. With this in mind, we aim to theoretically contrast the hypothesis that the modular character of food-sharing networks, may be an optimal response for avoiding the risk of individual and collective starvation.

Characteristics and drivers of food-sharing networks. It has been reported that empirical food-sharing networks present the same traits of other general social networks such as *degree assortativity* [11], or positive correlation on neighbors degree, *transitivity* or *global clustering* [284], the fact that agent i is likely connected to k if i is connected to j and j to k. These two network characteristics may be an indication of communities [212]. Other such characteristics are *reciprocity* [169], which is a significant proportion of agent pairs with bidirectional connections, and a *multilevel* structure [89] of cohesive groups hierarchically organized. Among the few theories aiming to explain some of these network patterns, it has been argued that the multilevel structure would be related to the efficiency of information flow [200] which may in turn, facilitate the diffusion of cultural innovations [80]. We think that, though the efficiency information flow may be an important force shaping social networks, it does not fully serve as an explanation for the evolution of food-sharing networks. On the one hand, it does not have a clear connection to the usual motives for food-sharing discussed in the evolutionary anthropology literature [124]. On the other hand, the theory of cultural innovation may impose high cognitive requirements [81] which do not explain, for example, why these multilevel networks may be observed also in other social animals and non-mammals [56]. Another work [89], though it is not a theory of multilevel structure, gives suggestive evidence relating three nested network levels, the household, the cluster, and the camp, to distinct social and economic cooperative functions: intersexual provisioning of a couple and their dependent children in a household, kin provisioning, and risk reduction reciprocity, respectively.

The motives for food-sharing most supported by empirical field studies are kinship, need, reciprocity, and inter-household distance, where the latter is usually considered a proxy of the tolerated theft motive [214, 169, 158, 155, 146, 128, 8, 257]. The costly signaling hypothesis, which claims that food-sharing is made in order to gain status, has received minor support in recent works as a factor for food transfers, though status is an important driver of reproductive fitness [283, 121, 213]. Probably the most accepted causal interplay between these motives is the following. Reciprocal altruism, or contingency on past transfers, arises as an important, or the strongest, predictor in most studies, usually explained as a

means of lowering the uncertainty of food generation [89, 257, 8, 128, 146, 155, 158, 214, 169, 222]. Kinship, or the preference for relatives, is another important driver of food transfers [169, 214, 158, 146, 8, 89, 257, 222], which appears to be as a bias, or partner-selection mechanism, interacting with reciprocity or need by increasing the associations to transfers with respect to non-kin [8, 214, 169, 158]. Tolerated theft, or sharing when the benefits of hoarding are outweighed by the cost of retaining the resource, is often a less important but significant motive of transfers [146, 123], though it may be empirically difficult to disentangle from the alternative explanations [123, 222, 146]. Some works have suggested that tolerated theft may be an evolutionary precursor of reciprocal altruism, given that the former is more common among other animals, and the latter has higher cognitive requirements that makes it almost nonexistent in non-human animals [265, 145]. Finally, the motive of need, or giving to those with a lower relative net energy production, often reflected in a positive correlation of transfers with receiving household size, is also an important force guiding food transfers [169, 158, 8, 257, 89].

The issue with need motive and a possible solution. The motive of need for food transfers has a difference with the other reviewed motives: it is not evolutionary adaptative, at least in the usual individual fitness models [151]. These models assert that cooperation requires to be assortative to be stable, that is, directed to other cooperative individuals [33]. This prediction has only partial support in real-world food sharing, where it is common that families or foragers in relative need like the elderly, tend to receive more resources [89, 8, 257] or that free-riders, those who take shares, but do not reciprocate, are not excluded from sharing [34]. A possible theoretical remedy that has been suggested [257] is to consider a diminishing returns value function of the transfers, which may produce a long-term equilibrium in transfers. In a previous publication [120], it was argued that assuming reciprocal altruism, sharing is an optimal individual strategy for values following a diminishing returns function of the shares, rather than just the raw quantity of food.

We propose a distinct intuition guiding the abstraction of our model. We propose that the argument for using diminishing returns comes from the fact that the essential problem a hunter-gatherer community is solving when sharing food, is to satisfy the need given by hunger. This is the intuition of the law of diminishing marginal utility given by Mill [106], stating that if there is enough of a good to satisfy a basic level of need, then each additional increment of the good satisfies the need less than the previous increment. The basic social dilemma for the food-sharing network then comes from the conflict between the goal of satisfying the own need and the need of other people. If the satisfaction of the hunger need is defined as the goal of minimizing the risk of starvation, or the time without eating, then the own need may be embedded in the tolerated theft or reciprocal altruism motives, while the hunger of other people may be represented by the need motive. We do not explicitly incorporate kinship in our model in order to keep it simple. It has been suggested that many kinship configurations may be a product of sharing patterns [180, 136, 152], which suggests that the basic social dilemma posed is enough to recover the basic dynamics of food-sharing networks, such as cohesive groups given by nuclear families with a few interconnections that comprise a cluster of akin families [89].

Our proposed model for food-sharing networks, its assumptions, and contributions. In this work, we propose a normative model [195] for food-sharing networks, formulated as a network optimization. That is, the model does not aim at fitting real networks, but rather derives idealized results from a set of simplified assumptions. The optimizing goals correspond to the minimization of starvation risk, either at the individual level or as a group. These goals are named in the model as *Reduction of Variability* (RV) and *Welfare* (WEF). The assumptions taken by the model are the following.

- An individual who has food eats only the amount required to survive, and shares all the remaining food. This is inspired by the empirical finding that giving a high proportion of food is a common practice [122].
- The structural imbalance between production and consumption leading to adult sharing to provide infants and juveniles [124], is implemented by a special set of network nodes called *hunters*, which are those who inject food into the network. The number of hunters in a network of N agents corresponds to the parameter nh.
- Since large and unpredictable food items such as meat or honey are more widely shared [122, 145, 34, 130], the model poses that at each time step, there is an amount F of food able to feed several individuals that is produced with certain probability ph. These quantities, F and ph, are parameters of the model.
- Any additional motives for sharing, like kinship, which would introduce, for example, preferential sharing partners, are abstracted. Therefore, given a food-sharing network, and an agent that receives food, this agent chooses randomly one of their neighbors to share. Thus, the transfers of food in the model correspond to recurring exchanges among stable connections, that may well reflect the food sharing in a small community formed by a few akin families. This assumption is also consistent to the choice taken for sampling the population size parameter used in the simulation. Technologies for storing food are not considered in this model. These assumptions are taken to keep the model simple.

The model aims to recover the constraints for optimal network organization derived from material transfers of food under the stated assumptions. This emphasis in the material dimension of food-sharing, does not look for neglecting that this behavior is a complex sociocultural phenomenon [3]. Rather, the model shows that evolutionary models of social dynamics may benefit from a resource distribution perspective and the explicit inclusion of the group level of analysis, given that the typical small-world structure of social networks given by the presence of communities, is recovered under the simultaneous optimization of the two goals, that is, the individual but also the group survival. Specifically, the contributions of our work are the following.

First, networks optimizing both criteria may exhibit a community structure of cohesive groups under stringent conditions of food supply. These networks are resemblant to the food-sharing networks observed by Dyble et al (2016) [89], where there are cohesive groups corresponding to households provisioned by an adult couple, and a set of households forms a cluster with a small number of inter-household connections. This model result suggests that this organization of nuclear families may have been evolutionarily functional for resource

distribution optimality, and can be considered a new mechanism for the emergence of communities in networks, as those reviewed in Section 2.4 from the survey chapter. We have additionally obtained that optimal welfare networks, where each hunter is connected to one big, homogeneous and dense group of non-hunters, resemble social networks that promote more egalitarian income distribution in a lab gift game reported by Chiang (2015) [64]. This model result is consistent with the view that social networks structured originally according to ecological considerations, may have created the environment in which prosocial tendencies and equity response elicitation evolved [144, 139]. Finally, distinct distributions of reciprocity are obtained for each optimization regime, which may be consistent with the empirical finding [153] that often the sharing occurs in waves: first among hunters, and then hunters with their families. This model result gives a broader picture of the usual notion that reciprocity is driven by the minimization of food production uncertainty [155].

Chapter structure. The rest of the chapter is structured as follows. In the next section, "Materials and methods" (Section 4.2), all the modeling assumptions and analysis techniques are described at a high level. Several details are set aside for the Annex D of supporting information. Then, in Section 4.3 the results of the analysis are shown, where a description is given regarding the structural behaviors observed in the three types of optimal networks analyzed, WEF, RV, and simultaneous optimization of both criteria. The interpretation of these results is addressed next in Section 4.4. We end the article with the conclusions.

4.2 Materials and methods

This section aims to briefly explain the modeling, implementation and analysis approach adopted in this work. Several details are left out to be consulted in the Annex D. The next sections go as follows. In the first two sections we describe the abstract protocol of food-sharing among agents, and the consequent model of probabilities of receiving food. Next section describes how the optimizing goals, welfare, and reduction of risk, are designed as a function of these probabilities. Next, Section 4.2.4 shows the solution concept of the optimization model, and how this model is implemented by evolutionary algorithms. The section is finalized by the pipeline of analysis, which describes how the output of the optimization model is processed using the following data analysis techniques: description by network features, clustering, and classification trees.

4.2.1 Food-sharing protocol and assumptions

We propose a network optimization model in which each agent is a node in a loop-free directed network $D = (V, E)$. There is a special set $H \subseteq V$ of agents in D that we call *hunters*. Agents in H are, intuitively, responsible for injecting *food* in the network by *hunting preys*. We assume that there is a real number $ph \in [0, 1]$, a parameter of the model, that determines the probability that an agent in H hunts at each time step. The prey obtained in every hunt is enough to feed F agents (including the hunter). The value F is a fixed positive integer that is also a parameter of the model. The food resulting from a hunt is *shared* from the

hunters to the rest of the network, only through network edges $e \in E$, and is used to feed the agents as we next explain. These assumptions implement the following two empirical facts: food items more widely shared are those unpredictable and coming in large packages [122, 145, 34, 130], and the fact that all agents require eating for survival, but only some of them can produce food [124].

We assume that all the exchanges of food derived from one successful hunt, occur at the same time step where the prey is caught. When an agent $v \in V$ receives f units of food from a neighbor, it consumes a single unit of food. After this, agent v chooses uniformly at random one of its neighbors, say v', and sends all the remaining food ($f - 1$ units) to v'. This assumption is inspired by the empirical finding that giving a high proportion of food is a common practice [122]. Notice that this sharing from v to v' is possible only if v initially received $f > 1$ units of food. Additional food received in the same time step, cannot be mixed with other food packages, nor stored for later usage. For simplicity, we assume that a hunt by a hunter is equivalent to receiving F units of food, though in this case the food is not received from a neighbor but from the environment. Thus, after feeding itself, a hunter sends $F - 1$ units of food to a random neighbor. These assumptions are taken to keep the model simple and to abstract the possible motives for agents' actions.

In our model, time is discrete, and at each time step, every hunter may or may not hunt with the same probability ph. Due to previous assumptions, those agents at distance greater than $(F - 1)$ to every hunter, simply cannot receive food. Thus, every node $v \in V$ in the network at a distance at most $(F - 1)$ to some hunter, may or may not eat depending on whether v receives food, either by hunting or by sharing. In order to model the effect of feeding an agent, we assume that agents have a *life span* of $n \in \mathbb{N}$, and that there is a *critical time window* $k < n$, in which, a run of k consecutive time steps without receiving food turns out to be seriously detrimental for every agent, resulting in its death. Only non-dead agents can hunt, eat, send or receive food. Therefore, the agents will try to eat as frequently as possible, to avoid spending too many rounds without eating.

To sum up, the parameters of our model are the probability of hunting in a time step (ph), the units of food produced by a prey (F), the number of time steps of life span (n), and the maximum number of time steps an agent can survive without eating (k). Moreover, a network in our model is defined by $D = (V, E)$ plus a set $H \subseteq V$ of hunters. We note that, even when fixing the parameters of the model, different networks behave in distinct ways depending on their connectivity (and the presence of hunters).

4.2.2 Probability of eating

In this section, we give some intuition about how the probabilities of eating behave according to the previously described protocol of food sharing. The interested reader may look at Annex D.1 in order to see the analytical derivation of these probabilities. For now, let us see an example. Figure 4.1 shows a network with 2 hunters, 0 and 4. Assume that $ph = 0.2$ and obtain the probabilities of eating according to (D.1) for distinct values of F, displayed in Table 4.1. For $F = 2$, 5 is surely fed by 4 with the probability of hunting ph, while 1 and 3 get a distributed chance of receiving food. In the case of $F = 3$, now 6 is also fed with

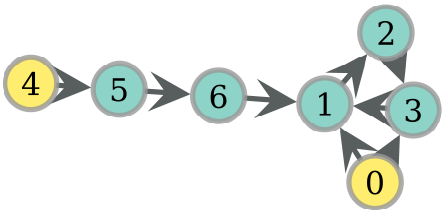


Figure 4.1: **A network with 7 nodes and 2 (yellow) hunters.**

Table 4.1: **Probabilities of eating, distinct F values.**

F \ $pe(\cdot)$	0	1	2	3	4	5	6
2	0.2	0.1	0	0.1	0.2	0.2	0
3	0.2	0.2	0.1	0.1	0.2	0.2	0.2
4	0.2	0.36	0.2	0.2	0.2	0.2	0.2
5	0.2	0.36	0.36	0.2	0.2	0.2	0.2

probability 0.2, and there are two 2−walks from 0: $\{(0,1),(1,2)\}$ and $\{(0,3),(3,1)\}$. Since 1 is fed by the 2 equiprobable walks, its probability of eating is $2 * 0.1$. When $F = 4$, 1 gets food by another walk, raising this probability to $0.36 = 0.2 + 0.2 - (0.2)^2$ thanks to Inclusion-Exclusion. Finally, notice in $F = 5$ that, though node 3 is reached twice by one walk, this does not increase the probability of eating since this probability does not measure the quantity of food, but rather the distinct events of food supply.

4.2.3 Optimizing criteria

We have said our model may have two possible goals, welfare, the egalitarian group access to food, and avoidance of individual starvation. Since our model implements food exchanges as a probabilistic protocol, the access to food resources is expressed as the probability of eating. Our model is framed as the minimization of an objective function, thus, we aim to penalize the dispersion of this probability among agents. Next, we formalize the two goals as minimization objectives.

Definition 1 (Welfare) Given fixed values for variables ph, F, n, k and $nh = |H|$, *Welfare* (WEF) criterion for minimization is the standard deviation of probabilities of eating over network nodes.

Now, the minimization of individual starvation risk is formalized as follows. We can think of the time steps in our model, and the possibility of eating at each time as a sequence of Bernouilli trials, where a success is given by the event of not receiving food, otherwise we have a failure. That is, at each time step, node v obtains a success with probability $1 - pe(v)$, where $pe(v)$ is computed as explained in the previous sections. Since network nodes attempt to avoid k consecutive steps without food, or a success run of length k, in order to stay alive, a possible formulation of reduction of starvation risk can be expressed as the minimization of the number $G_{n,k}$ of success runs of length greater than or equal to k in n trials. The distribution and moments of this random variable can be computed exactly with a Markov chain embedding technique [105], which allows defining this criterion as the minimization of the sum of the expected value and standard deviation of $G_{n,k}$.

The random variable $G_{n,k}$ defined above is expressed as an integer number of runs. One can normalize this variable $G_{n,k}$ by the total maximum number of runs of length k that may be observed in n trials. In this way, this criterion could be interpreted as the probability of dying within the life span -or observing a k−length (or greater) run with no food in n

trials- as a function of the probability of receiving food at each trial. Since the number of runs of length at least k within n trials is maximized when all the runs have the minimum length k, the normalizing denominator may be estimated as the floor of $\frac{n+1}{k+1}$, where the "+1" appears to count the spaces necessary to split a new run from the previous one. With these ingredients, we formally define the second minimization criterion.

Definition 2 (Reduction of variability function) Given fixed values for variables n, k, and $1-pe(v)$ for the probability of a success, let $G_{n,k}$ be the random variable given by the number of success runs of length greater than or equal to k in n trials. We will refer to the expected value and the standard deviation of this variable as $\mathbb{E}(G_{n,k}(pe(v)))$ and $\sigma(G_{n,k}(pe(v)))$. The *Reduction of Variability* of food intake (RV) function of the probability of eating, is defined as $RV(\cdot) = \frac{k+1}{n+1} \cdot (\mathbb{E}(G_{n,k}(\cdot)) + \sigma(G_{n,k}(\cdot)))$.

Definition 3 (Individual starvation risk criterion) Given fixed values for variables ph, F, n, k and $nh = |H|$, the *Individual starvation risk* criterion for minimization is the average over network nodes of the RV function, evaluated at the respective probability of eating of each node. For brevity, and without causing ambiguity (since the function is for nodes, and the criterion for networks), we refer also to this criterion as RV.

We have taken the average to approximate the individual agent risk since this function is not a robust statistic of central tendency [248], and may be extremely affected by the presence of outliers. That is, the mean of RV over network nodes may be minimized by a few number of agents with low RV, while a great number still experiment high starvation risk. Now, in Figure 4.2 we can observe distinct instances of the RV function, for different values of k and $n = 1000$. These RV curves have a maximum value corresponding to some value $pe^* \in (0, 0.5)$, and the larger the length k of the run, the smaller the probability of eating pe^* at which this maximum is obtained, which makes intuitive sense. In the following, we will continue working with the values $k = 10$ and $n = 1000$, since from inspection of this plot, this function displays all relevant behaviors within a non-degenerate probability domain. Therefore, the resting free variables, with which we will work are F, ph and $nh = |H|$.

4.2.4 Model solution

Definitions. We need to define the set of solutions for our minimization model. At first sight, this set corresponds to the directed graphs of N nodes, if we have a population of N agents. However, there is a huge number of non-isomorphic networks sharing the same optimizing costs due to subgraphs in the network that do not contribute to the optimizing criteria functions. We remedy this by defining the domain set of our optimization model as the quotient set of the set of networks of size N, with respect to an equivalence relation given by a suitable definition of isomorphism that accounts for this fact. This definition impacts the evolutionary algorithms used to compute model optima, which are described in the next section. The interested reader may review the isomorphism definition and examples in the Annex D.4.

Finally, we give the solution concepts used in our minimization problems. For single criterion minimization, a solution network is given by a *Local Minimum*, which is a network of

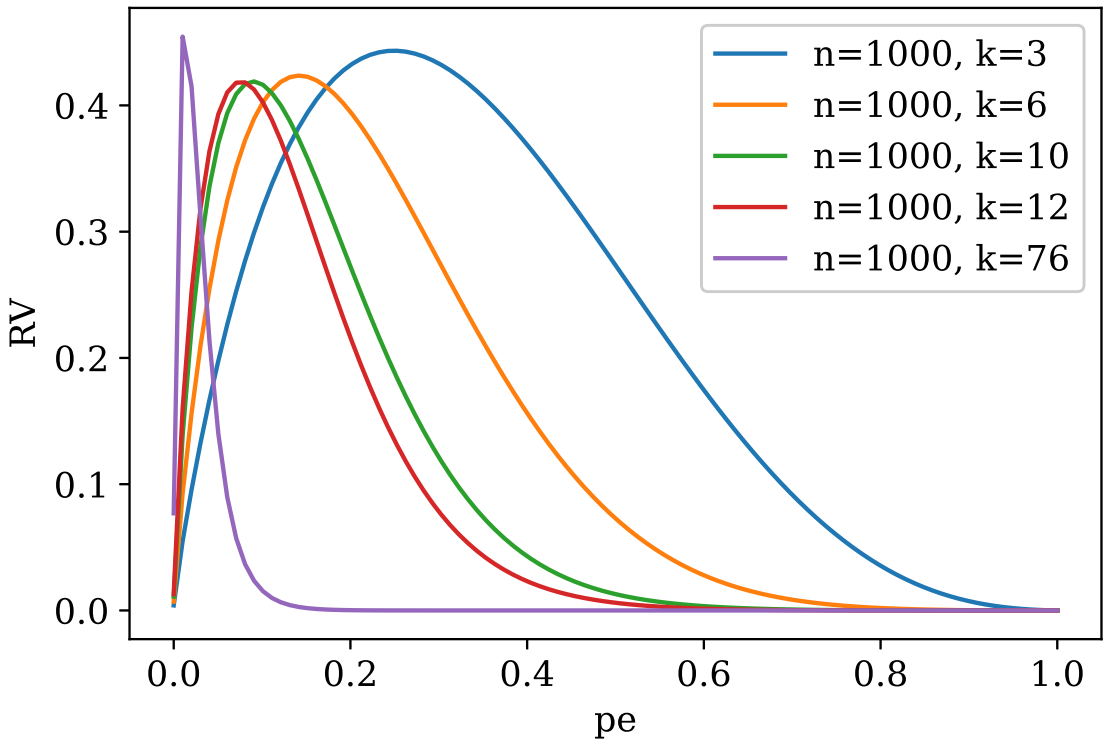


Figure 4.2: **Reduction of variability function of the probability of eating.** The X-axis is the probability of eating or receiving food, as defined in the last sections. The vertical axis is the Reduction of variability (RV) function of this probability. Each curve corresponds to a distinct instance of the variables n, the number of life span time steps, and k, the maximum number of time steps an agent can survive without eating. Notice that, as the survival time without eating (k) gets smaller, the probabilities of dying (RV) are greater for a wider range of food availability (pe), which engenders more 'fragile' societies, with small or absence of food availability for non-hunters (see Section 4.3.2).

minimal value with respect to every possible change of network arc. In the case of simultaneously minimizing both criteria, we use the standard notion of *Pareto Optimality* [69], defined next.

Definition 4 (Dominance) A vector $\vec{u} = (u_1, u_2) \in \mathbb{R}^2$ is said to *dominate* vector $\vec{v} = (v_1, v_2) \in \mathbb{R}^2$, which is denoted by $\vec{u} \preceq \vec{v}$, if $\forall i \in \{1, 2\}$, $u_i \leq v_i$, and $\exists i \in \{1, 2\}$, such that $u_i < v_i$.

Definition 5 (Pareto Optimality) Let $F(\cdot) = (f_1(\cdot), f_2(\cdot))$ be a set of objective real-valued functions. A point x from the domain of F is said to be *Pareto Optimal*, if there is no x' in the domain of F, such that $F(x')$ dominates $F(x)$.

Defined in this way, a point $x^\star$ is Pareto optimal for a given objective function $F(\cdot) = (f_1(\cdot), f_2(\cdot))$, if there is no feasible point x, such that $f_1(x) < f_1(x^\star)$ and simultaneously, $f_2(x) < f_2(x^\star)$. The set of all Pareto optima is the *Pareto optimal set*, and its image via function F is the *Pareto front*. Now, the computational time complexity of finding the Pareto optimal set, from a total set of two-dimensional vectors of size N, is given by a number of $O(N \cdot \log N)$ comparisons [177]. Since our total set given by the food-sharing networks of size 12 (see Annex D.2) has an intractable size, an approach alternative to the procedures by Kung et al (1975) [177] must be taken. In the next paragraph, we describe the meta-heuristic solution we have implemented for this problem.

Model Implementation. We have tackled our minimization problem with evolutionary algorithms, due to their versatility to accommodate the model assumptions, and their ability to manage several candidate solutions. We have used the known $(\mu, \lambda)-$procedure, where μ is the population size, at each generation every individual gives birth to λ offspring by a feasible mutation, and from this set μ descendant are selected for the next generation population. This procedure is recommended as a way to preserve the diversity of solutions [91]. For single objective offspring selection, we have used *tournament selection* of size $t = 12$, that is, t individuals are randomly selected from the offspring population, and the one with the best function value is selected. This relatively high value of t is chosen to increase the likelihood of selecting local optima, something usually referred to as *exploitation* or *selective pressure*. In most cases, we have used $\mu = 1000$ and $\lambda/\mu = 8$ in order to increase the selective pressure. The multi-objective case employs the NSGA-IIR algorithm [102] as an offspring selection mechanism. This algorithm is an improved version of the well-known NSGA-II [79], that takes advantage of its relatively low computing and storage complexity, but that solves some instabilities that may appear when two or more solutions share the same objective values [102]. On the other hand, for two optimization objectives and a population size μ, the time complexity of this algorithm is $O(\mu \cdot \log \mu)$ in the average case and $O(\mu^2)$ in the worst case, while its space complexity is $O(\mu)$ [101]. We have implemented these algorithms on the DEAP library [100] for Python, a package devised for the modular design of evolutionary algorithms. The networks are implemented with the module graph-tool [226], a Python library for manipulation and statistical analysis of graphs. We have taken a series of implementation decisions to obtain results more efficiently. The interested reader may review them in the Annex D.4.

4.2.5 Pipeline of analysis

The outcome of the evolutionary algorithms described previously is a set of networks, for every sampled setting of parameters F, ph, and $nh = |H|$. In the Annex D.2, the structure of the simulation carried out is explained, which comprises the choice of sampled parameters. Now, to obtain a high-level description of the set of different optima networks, we processed the data according to the following steps. For each network, a set of features is computed to build a table of networks and features, for either single criteria or multiobjective case. See Tables D.1, D.2 in the Annex D.3 to consult the sizes of these datasets. These networks are then clustered by their feature similarity, where each group depicts a set of qualitatively similar networks. Finally, a decision tree to discriminate distinct groups is computed, which permits an objective split of groups in terms of the feature values. Next, we review in more detail each of these steps.

Descriptive features. Here we explain the features used to describe the networks. These features correspond mostly to simple metrics of network topology that can be efficiently obtained. We next list them grouped by their reason to be included as potentially useful descriptors. An important remark is that, though these descriptors are used in their original scale to build the decision trees, they are standardized for the clustering stage since this is recommendable for these methods. We used the generalized spatial sign technique to accomplish this [240], with quadratic radial function and $k-$step least trimmed squares for the location estimator, since it delivers robustness to estimators based on co-variances in the presence of potential outliers.

- The minimization of WEF may reduce the dispersion in the number of walks from hunters to every node, since the probabilities of eating depend on them. We aim to detect this scenario with the proxy variables *standard deviation of out-degree and in-degree distributions*.

- To identify potential community or group structure, we have added several variables. A very simple measure of this is the *number of strongly connected components*. We added the mean of *local clustering* [286], which estimates the probability that two neighbors are connected, and the *in-degree and out-degree correlation or assortativity*, since they have been associated with the presence of dense groups [212]. Assortativity has also been associated to the presence of *reciprocity* [296], which we define as the proportion of connected vertex pairs having bidirectional arcs. This is a prevalent pattern found in social networks [285]. We have finally considered *network modularity* [211], an index of network partitioning into dense groups with few inter-group connections. We obtain candidate optimal partitions by running a Markov chain Monte Carlo method several times [225].

- The minimization of RV may produce important variations in the overall connectivity of the network, depending on the specific value of ph. We will measure network connectivity with the *mean of out-degree distribution*, and the *mean of in-degree of hunters*.

The latter variable may also diminish in some WEF minima.

- The 2 network costs to be minimized, RV and WEF.

The assortativity indexes have been used only for WEF analysis, since the other type of optima sometimes presents indefinite values.

Feature-Similarity-based Clustering. The stage of clustering is performed by the successive application of tSNE and OPTICS algorithms. The first of them, the *t-Distributed Stochastic Neighbor Embedding* [275], was chosen by its formulation which finds a low-dimensional map preserving the local similarity structure of the data. This implements one of the goals of our analysis which is to group similar networks, where their similarity is expressed in terms of their feature values. Another important reason to choose this technique is its remarkable ability to produce reliable low-dimensional visualizations of big datasets, which allows us to perform an intuitive assessment of the quality of the solution. The interested reader may review in the Annex D.5 the election of hyperparameters of tSNE. The second part of the clustering, OPTICS, or *Ordering Points to Identify the Clustering Structure* [10] is used as an unsupervised method to discriminate the distinct groups produced by tSNE. OPTICS is a density-based clustering method where clusters are regions in which the objects are dense and are separated by *noise* or regions of low object density, making these approaches to be flexible enough to accommodate clusters of arbitrary shapes and distribution of points. We decided to use tSNE plus OPTICS, instead of only OPTICS, since we empirically checked that the former methodology resulted in better identifiability of clusters of similar networks. Now, the hyperparameters of OPTICS are usually dependent on the dataset, so there is no direct way of setting them. Our approach was to devise heuristics to choose these parameters. The classes produced by OPTICS with this selection of parameters visually match the dense regions of tSNE with coarser separation. The interested reader may review the Annex D.5 and Table D.3 for details on the heuristics.

Description of clusters by classification trees. After obtaining a robust clustering of each dataset, we compute a series of decision trees (DT) to classify the networks into the distinct groups produced by the clustering stage, trained with the same features employed in the clustering, plus the model variables F, ph, nh, $F * nh$, $ph * F * nh$ and the mean of the probability of eating for non-hunters. In the Annex D.6, the interested reader may review an informal argument for the importance of these features. Now, the classification tree model we use corresponds to that proposed by Loh (2009) [189], since it offers a robust split selection strategy based on selecting the variable with the most significant chi-squared main effect, which redounds in predictive accuracy and model parsimony. Since we look for a robust characterization of the dataset clustering, we compute several trees, each trained with a bootstrap sampling with replacement of size 70% of the dataset, thus it is expected that each training set has approximately half of the set of unique points of the dataset [90]. These computed trees are examined to find the most reliable ones, which we define as those trees that are Pareto optimal for maximization of balanced accuracy [125], which accounts for possible imbalanced classes, and a measure for syntactic stability. We are interested in

syntactic stability due to the fact we use the DT representation to draw conclusions about the dataset. Several details regarding the validation of the decision trees may be consulted in the Annex D.7. Decision trees hyperparameters used can be consulted in Table D.4.

4.3 Results

In this section, we report the analysis of model optimal networks, carried out through the pipeline of analysis described in the last section. We review the typical structure of the three types of optima, RV, WEF local optima, and Pareto optimal networks for both criteria, with an emphasis on the behavior of some patterns found in social networks, namely, reciprocity and community structure. In this sense, we have found WEF optima display high clustering and low network segmentation, with no distinguishable modules. Local optima for RV present more diversity, ranging from cooperation exclusive among hunters to networks similar to WEF optima. They may present community structure, but with rather sparse connectivity. Finally, Pareto optimal networks show even more diversity, including either networks similar to the already described local optima, but also new configurations such as distinct cohesive modules around hunters, or high reciprocity between a hunter and non-hunters.

4.3.1 Welfare optima

Summary of distinct types of WEF optima. It is instructive to start by inspecting the following tree of WEF optima in Figure 4.3, belonging to the accuracy-stability-Pareto optimal set, which has an average accuracy of 0.872 to discriminate the 7 clustering classes. In these trees, the loss corresponds to the Gini coefficient of samples at that node, the predicted class is the class with a maximum number of samples, and proportion is the relative proportion of node samples with respect to the total in the root. This total and the other statistics present in the tree are computed on the training set. The accuracy, on the other hand, is computed on the test set, which is the complement of the training set with respect to the whole WEF dataset of size 18686 (see Table D.1).

The first remark is the presence of the condition $F * nh \leq 13.5$ at root, which turns out to be very stable since in each of the 9 distinct trees found at the Pareto efficient tree set, the structure formed by internal nodes of IDs 0, 1, 3 is present. The threshold 13.5 is an artifact from averaging the two most proximal $F * nh$ values, 12 and 15, and since $F * nh$ is a product of two integer variables, the condition $F * nh \leq 13.5$ can be equivalently replaced by $F * nh \leq 12 = N$. A possible topological marker of model variable $F * nh$, which reflects the relative food abundance, is the mean in-degree of hunters, a variable with which there is a significant Spearman correlation of 0.81. Another possible marker is the number of strongly connected components (SCC), variable with which the correlation, in this case, is -0.574. These relationships are due to the fact that optimal WEF networks tend to form a densely connected group of non-hunters that is fed by hunters, except if there is enough food, in whose case there is only 1 strongly connected component where hunters are also fed back. An important exception to this trend in SCC is shown by networks in node 16, where there are only networks with $ph = 0.6$ and a mean of 3.796 SCC. That is, when there is enough

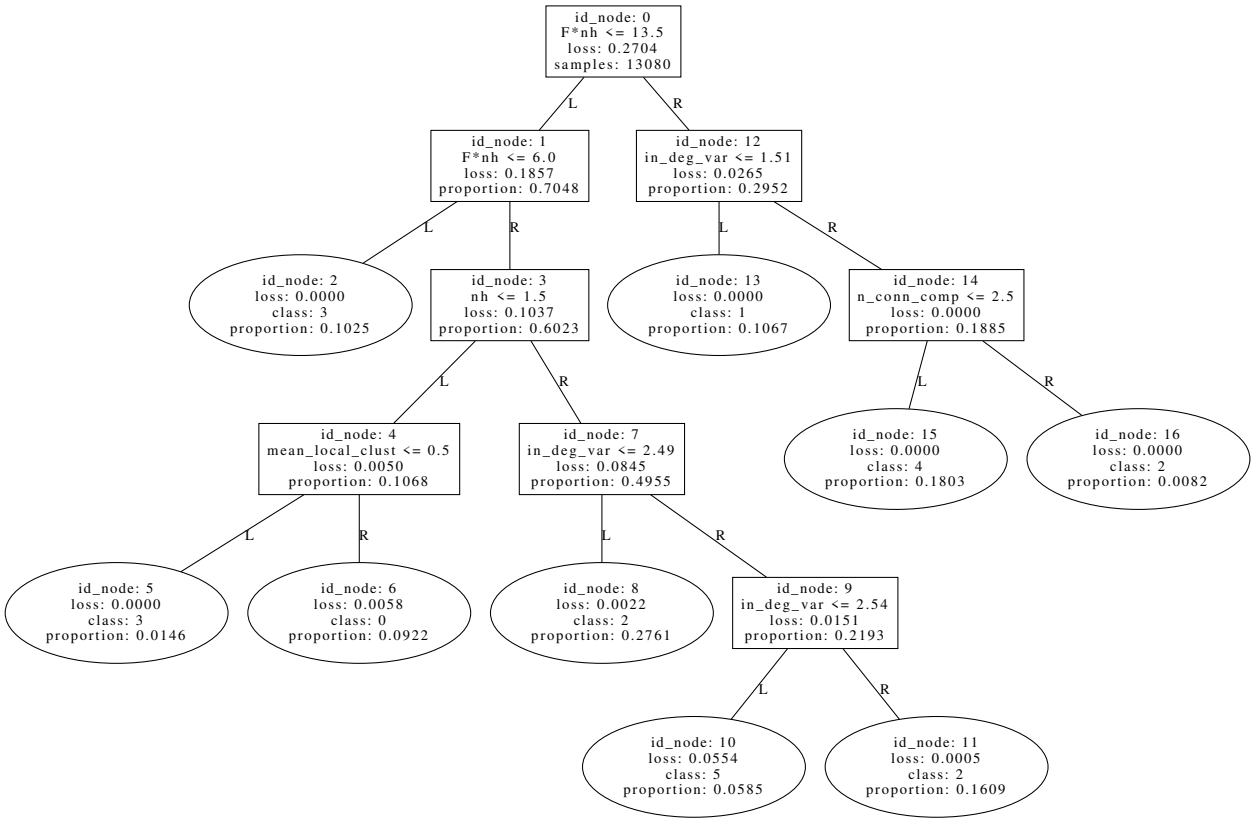
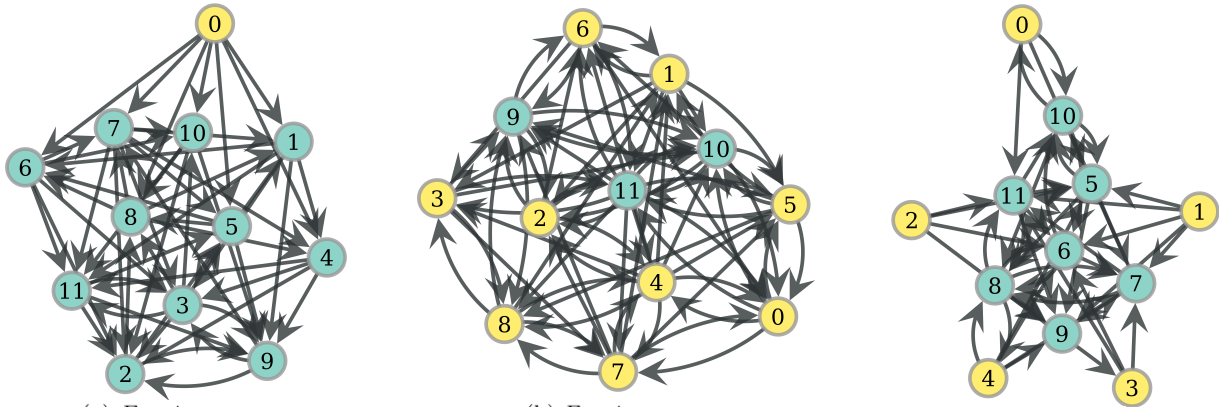


Figure 4.3: **An efficient tree (accuracy=0.872) to discriminate the clustering labels of Welfare optima.** Statistics in the tree are computed on the training set, while average accuracy is computed in the test set. See Annex D.3 for the sizes of the datasets used, and Paragraph *Description of clusters by classification trees* (Section 4.2.5) for the general procedure of tree construction. See the first paragraph from Section 4.3.1 for an explanation of the variables displayed in tree nodes. The left child subtree of the root corresponds to WEF networks where the food supply does not suffice to feed all the agents (see Annex D.6), while the right subtree represents networks with enough food, allowing more diversity in the structure of WEF networks, such as those in Figure 4.4 (b), (c).



(a) $F = 4$, $nh = 1, ph = 0.3$ (b) $F = 4$, $nh = 9, ph = 0.3$ (c) $F = 3$, $nh = 5, ph = 0.6$

Figure 4.4: **Most central networks of nodes** 2 **(left),** 13 **(center) and** 16 **(right) from tree in Figure 4.3.** Hunters are filled in yellow. Each network specifies the values of model parameters used as input, and for which the respective network is a local minimum. The structure of WEF networks seen in (a) with a densely connected group of non-hunters that is (externally) fed by hunters, appears to be necessary when the food supply is not enough to feed all the agents.

food with a high probability of hunting, the network is segmented into distinct SCC's that provide an egalitarian probability of eating. These structures are shown in the representative networks on Figure 4.4.

General statistics and modularity on WEF optima. WEF optima are networks of high clustering, mean out-degree, moderate reciprocity, and low assortativity and modularity, as shown in Table 4.2, which speaks of one large community with low segmentation. The WEF optimal networks with the largest modularity values, obtain this structure from the special feature of minimizing the standard deviation of in-degree. An example of these networks is in Figure 4.5 from node 8, having a modularity of 0.305 for the partition $\{\{0, 2, 6, 7, 9, 10\}, \{1, 3, 4, 5, 8, 11\}\}$.

4.3.2 Reduction of variability optima

Relation between RV and WEF costs. We start this section by analyzing the structure of the relation between costs RV and WEF, and derive from it a profile of the distinct networks obtained as RV local optima. We have mentioned in the Annex D.2 the intuition regarding the expected behavior of RV optima from RV function. That is, for values of ph smaller than $pe^\star$, the probability producing the RV maximum, there is an *every man for himself* regime, or free-riding situation where non-hunters just do not receive food and are isolated nodes. Now, for $ph > pe^\star$, there is *progressive cooperation*, an increasing trend of feeding non-hunters, particularly when RV reaches its range of flat values. On the other hand, depending on the values of model variables, hunters often share food in order to decrease the RV mean. We can visualize this pattern in Figure 4.6, where we graph the mean correlation values between RV and WEF for randomly sampled networks and distinct ph values.

Table 4.2: **WEF optima general statistics.**

Feature	Mean	St. dev.
Mean local clustering	0.481	0.167
Mean out-degree	4.763	1.173
Reciprocity	0.232	0.144
Out-assortativity	−0.082	0.044
Modularity	0.042	0.053

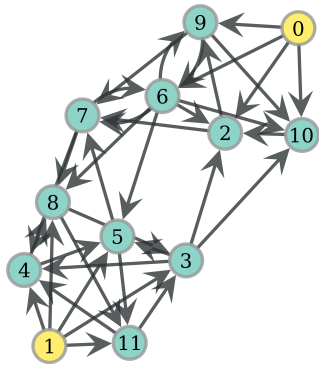


Figure 4.5: **A WEF optima minimizing in-degree variability.** Obtained under variables: $F = 4, nh = 2, ph = 0.6$. Hunters filled in yellow. WEF networks with not enough food for every agent, and several hunters, may have a modular structure with minimal in-degree variability (see Figure 4.3).

In this graph, it can be seen that these mean correlation values start approximately at -1 for low ph values, which corresponds to the free-riding regime, and in the extent ph increases the correlation curves tend to stabilize around values near to perfect correlation. Observe that most curves become increasing functions only for values of ph greater than $pe^{\star} = 0.0899$. The fact that greater values of the product $F \cdot nh$ have a faster convergence to cooperation may be explained, if one assumes a uniform feeding probability, from Equation (D.3), where a greater $(F \cdot nh)$ implies greater mean probability of eating, which makes these probabilities closer to the flat RV region.

Distribution of arc types and representative RV optimal networks. This pattern of costs behavior is reflected also in the distribution of distinct types of arcs. If we separate arcs in three groups depending on whether they connect, without arc direction: two hunters, a hunter and a non-hunter, and two non-hunters, we obtain the following distribution of the number of arcs depicted in Figure 4.7. It can be seen in the figure that, in the two smallest ph values, non-hunters are isolated in RV optima, which is part of the so-called every-man-for-himself phase. If ph increases, the relative numbers of arcs gradually converge to the same ordering displayed by WEF optima, which occurs for lower values of ph if $F \cdot nh > 12$. This pattern is consistent to what is displayed in Figure 4.6.

The distinctive profiles of the number of arcs from these graphs may be used to identify representative networks obtained as RV optima. In Figure 4.8, a decision tree that discriminates the classes of RV optima with an accuracy of 0.907 is displayed, together with representative networks from some of its leaves in Figure 4.9. In the left part of the tree, there are the networks with a lesser number of strongly connected components, which are networks with greater values of $F \cdot nh$. The networks from Figure 4.9 (a), (b), (c), show the prototypical behaviors shown in Figure 4.7 (c), which are, respectively, the absence of connectivity for non-hunters if ph is low, a network with WEF-like topology of minimization of in-degree variability if $ph \geq 0.3$, and a transition between the two regimes similar to the

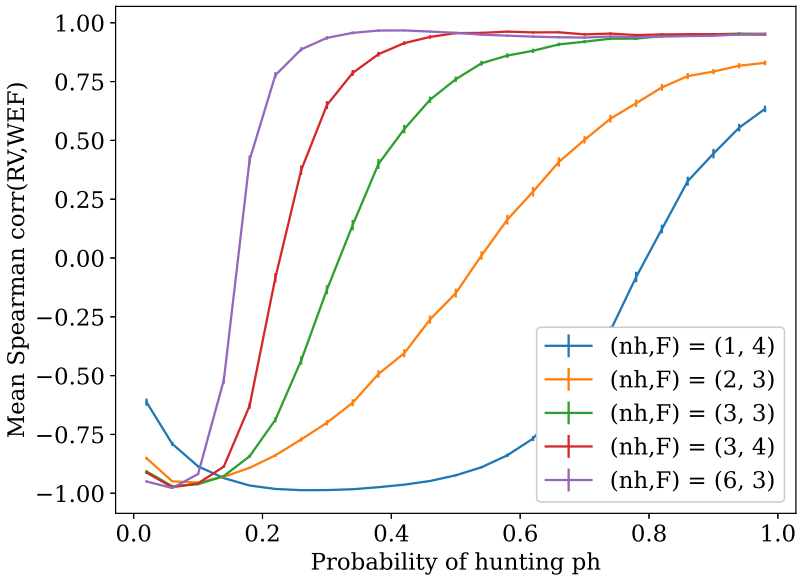
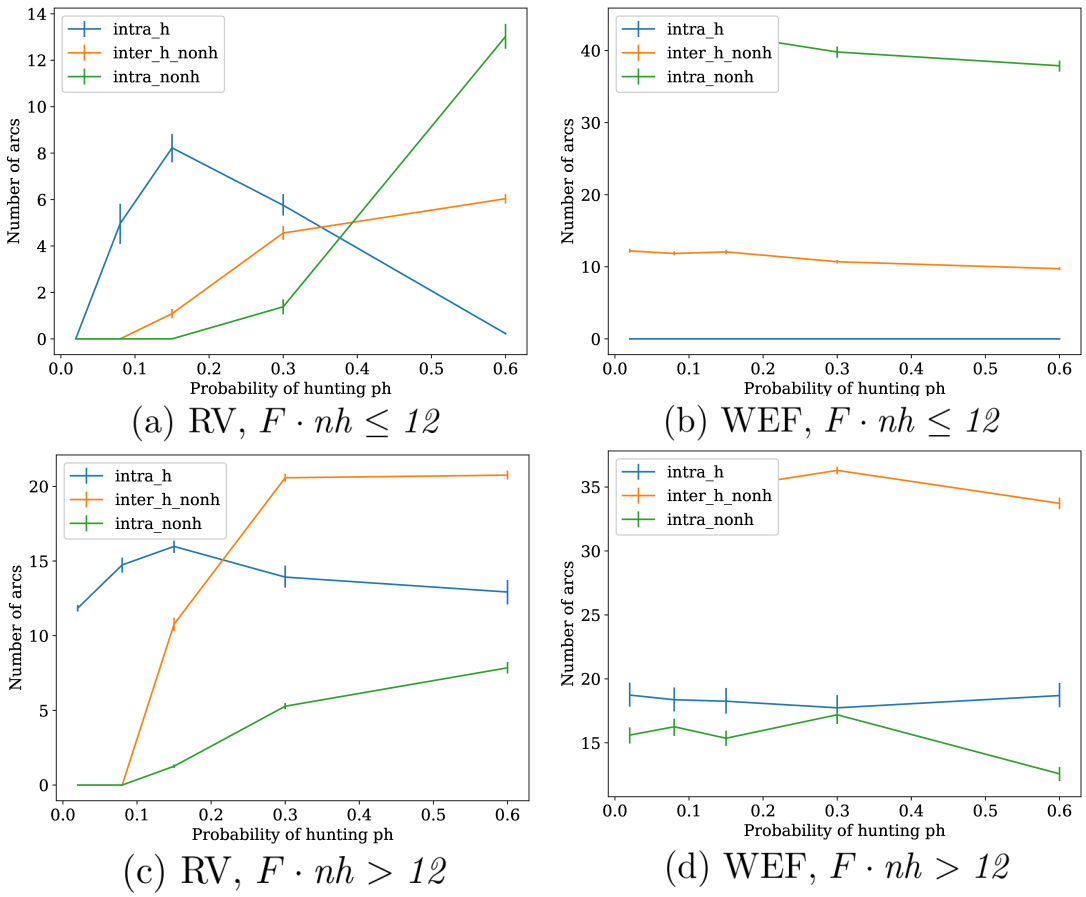


Figure 4.6: **Correlation of costs in function of the probability of hunting.** Each curve represents the mean of Spearman correlation between RV and WEF for a specific configuration of F and nh, and a range of 25 equidistant ph values. Each point is the mean of 100 correlation points, each one computed for a uniform random sample of 100 networks. Each mean point is surrounded by the percentile confidence interval using the standard error of the mean with error $\alpha = 0.05$, that is, assuming the sample mean is normally distributed. Observe that, for every curve, the average correlation starts to be increasing on ph from a given point, because the probability of dying (in terms of the RV function, see Figure 4.2) also starts to decrease with enough probability of eating, and this is valid for every agent. Also, with greater food supply ($F \cdot nh$, or greater probability of eating, see Annex D.6), for any given ph sufficiently large there are greater chances of survival for all the agents, which explains the distinct curve shapes.



(a) RV, $F \cdot nh \leq 12$ (b) WEF, $F \cdot nh \leq 12$

(c) RV, $F \cdot nh > 12$ (d) WEF, $F \cdot nh > 12$

Figure 4.7: **Distributions of arc type for distinct optima and food supply condition.** The three types of arc graphed (which do not consider arc direction) are: connecting two hunters (*intra_h*), connecting two non-hunters (*intra_nonh*), and connecting a hunter with a non-hunter (*inter_h_nonh*). Each number of arc type average is computed for every sampled *ph* value, and surrounded by a percentile bootstrap confidence interval of error $\alpha = 0.05$ and 1000 bootstrap samples of 70% of dataset. See the paragraph *Summary of distinct types of WEF optima* from Section 4.3.1, for the result justifying the use of condition $F \cdot nh \leq N$ to aggregate datasets. The plots show that with greater values of *ph*, the RV networks tend to look like WEF networks, and that this behavior appears for smaller *ph* values in the case $F \cdot nh > 12$, in accordance with Figure 4.6.

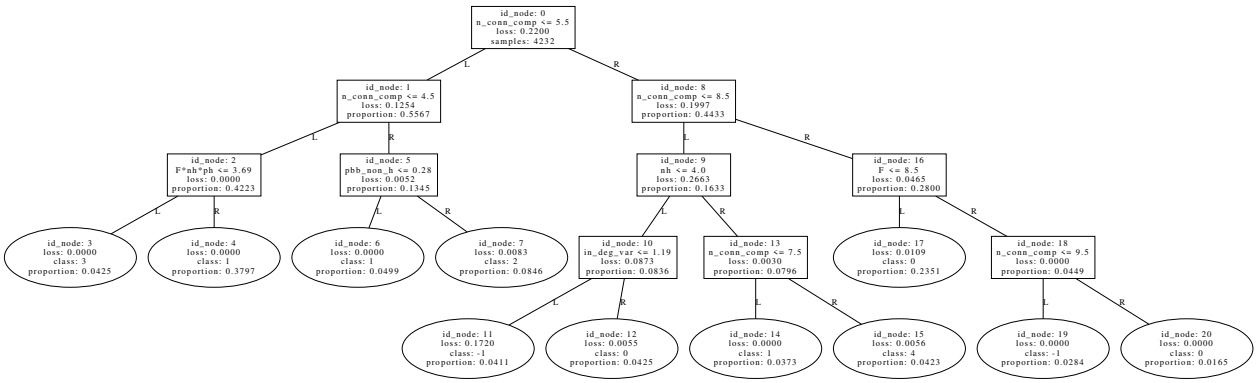


Figure 4.8: **An efficient tree (accuracy=0.907) to discriminate the clustering labels of RV optima.** Statistics in the tree are computed on the training set, while average accuracy is computed in the test set. See Annex D.3 for the sizes of the datasets used, and Paragraph *Description of clusters by classification trees* (Section 4.2.5) for the general procedure of tree construction. See the first paragraph from Section 4.3.1 for an explanation of the variables displayed in tree nodes. The main splitting variable in this tree corresponds to the number of strongly connected components. The left subtree has fewer components, like the networks (a), (b), (c) or (f) in Figure 4.9. Larger number of components are typically associated with more isolated agents.

last one, but where hunters have more intra-connectivity rather than inter-connectivity.

Now, some of the patterns from Figure 4.7 (a) are displayed in Figure 4.9 (d), (e), and (f). In the first of them, there is a network from Leaf 15, where networks have arcs mainly among hunters, and may present some few connections to non-hunters. The networks of this kind are among those in RV optima having the greatest proportions of reciprocated arc pairs, usually only between hunters. This is used as a strategy to lower the high RV values in this range of ph values ($ph = 0.15$). Then, in Figure 4.9 (e), there is more connectivity to non-hunters, but it is still comparable to that between hunters. Finally, Figure 4.9 (f) displays a network with the cooperative regime on $ph = 0.6$ where intra-non-hunters connectivity surpasses the number of connections to hunters, and these display a minimal intra-connectivity. The case when $ph = 0.6$ and there is just one hunter, as in Figure 4.9 (g), is special. These networks present a one-way walk with no forks, since a fork lowers the probability of eating with respect to the no-fork situation, increasing the RV cost. If there are isolated nodes, as is observed in most optima obtained from an evolutionary optimization, there is a big difference in the probability of eating between connected and isolated nodes, which increases WEF cost. This affects the Spearman correlation between ph and WEF in the whole dataset, which is 0.313. If we restrict to those data where $nh > 3$, this correlation lowers to -0.319, and to -0.73 if $nh > 6$, consistent with Figure 4.6.

Modularity on RV optima. We close this section by remarking on the behavior of modularity on this type of optima. In general, RV optima are networks of low modularity, see Table 4.3 for its general statistics, but in fact, high modularity may be observed. This high modularity is a by-product of survival or network arrangements for enhancement of probabilities of eating, as can be pointed from its correlation with $F \cdot nh$ which is -0.359.

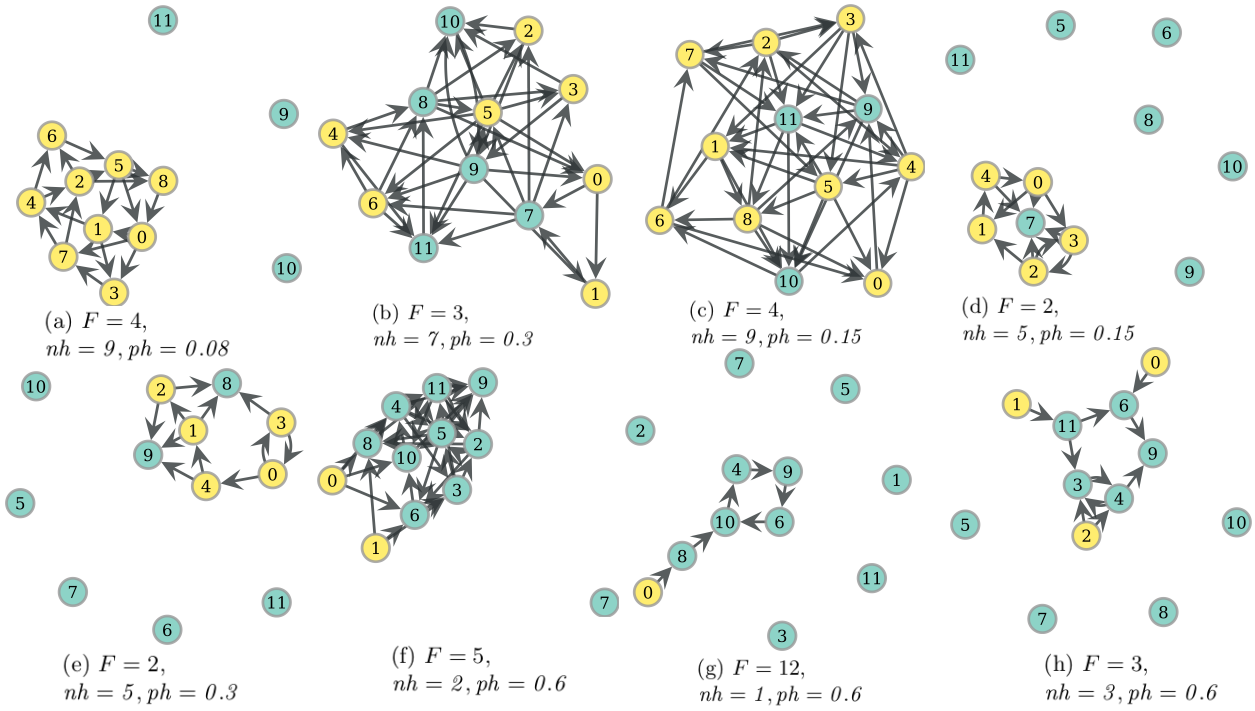


Figure 4.9: **Example networks from distinct DT leaves of the tree in Figure 4.8.** From left to right, and from top to bottom: (a) Leaf 3 (most central), (b) Leaf 4, (c) Leaf 4, (d) Leaf 15, (e) Leaf 17, (f) Leaf 7 (most central), (g) Leaf 19 (most central) and (h) Leaf 17. Each subcaption displays values of the condition under which these optima were obtained. Hunters filled in yellow. The networks show the diversity found on RV optima, ranging from networks like (a) with homophilic hunter links (see Section 2.4.4), to networks with an egalitarian (WEF) structure such as that in (f), or with modular structure such as (g), (h).

However, the relationship is stronger for higher ph values, since restricting to $ph > 0.2$ gives a correlation of -0.474. This higher modularity is usually more related to a segmentation in the network than to the presence of cohesive groups of nodes. An example of this is the same network from Figure 4.9 (g), which has a modularity of 0.208 for the partition $\{\{0, 1, 2, 3, 5, 7, 8, 11\}, \{4, 6, 9, 10\}\}$. Another example is that of Figure 4.9 (h) having a modularity of 0.32 for partition $\{\{0, 6, 8, 9\}, \{1, 10, 11\}, \{2, 3, 4, 5, 7\}\}$.

Table 4.3: **Modularity general statistics on RV optima.**

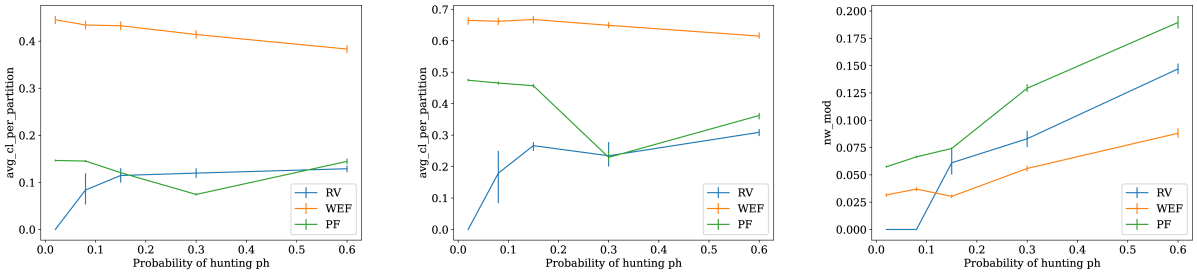
Percentile 0.5	Percentile 25	Median	Percentile 75	Percentile 99.5
0	0.029	0.08	0.145	0.48

4.3.3 Multi-objective optima

Overview and scope of this analysis. In this section we show some of the key social networks patterns that can be recovered in Pareto optimal networks from the simultaneous optimization of RV and WEF criteria, which we will refer to as PF (due to their criteria values being in the Pareto Front) optima. We start by showing graphs of average behaviors displayed by the distinct type of optima, highlighting the most relevant differences among them. We then illustrate these behaviors by depicting some representative Pareto optimal networks and conclude with a brief analysis. Since there is a wide spectrum of Pareto optimal networks, we restrict the study to the most interesting case of $F \cdot nh \leq N$. See the paragraph *Summary of distinct types of WEF optima* from Section 4.3.1, for the result justifying the use of this last condition. Most types of optima networks from the context $F \cdot nh > N$ appear also as optima in the former case. The interested reader may review the trees of the case $F \cdot nh > N$ and associated optimal networks, in the documentation for running the code from the repository provided for this purpose [233].

Community structure on PF optimal networks. One of the main results is displayed in Figure 4.10, where there are two measures of community structure, the already mentioned network modularity, and a new measure called *average intra-partition clustering*. This measure corresponds to the average of local clustering over the subnetworks induced by each of the modules of the partition maximizing modularity found by the Monte Carlo algorithm. Thus, intra-partition clustering is a measure of the cohesiveness of the partition maximizing modularity. A first remark to note is that, though maximum values of this measure are obtained by WEF optima, this pattern is more a product of the high connectivity across the whole network found in this type of optima, than to the presence of clear cohesive modules. This observation is supported by the low values of modularity exhibited by WEF optima, see Figure 4.10(c), and is coherent with the hypothesis that the partitioning found in this optima obeys more to the random nature of the Monte Carlo algorithm.

Now, if one inspects the mean values of average intra-partition clustering (AIC) of PF optima, one notes that they are not too distinct from those of RV (Figure 4.10(a)). As an example, for the observed PF networks with $ph \leq 0.15$, the mean is 0.148 and the median is 0.056, and one can find typical networks such as those in Figure 4.11 (a), (b), (c), with AIC



(a) Average intra-partition clustering.

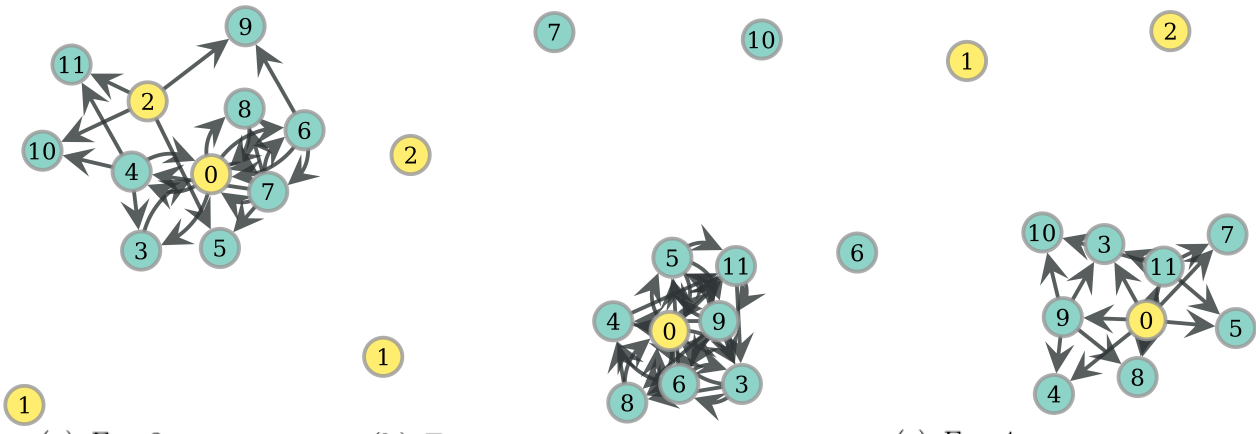
(b) Percentile 90 of average intra-partition clustering.

(c) Modularity.

Figure 4.10: **Distributions of different measures of community structure for each type of network optima.** From left to right: Average intra-partition clustering (left), percentile 90 of average intra-partition clustering. (center), and mean modularity (right), for each type of network optima. Average intra-partition clustering (AIC) is the mean of local clustering computed in each subnetwork induced by the modules of the partition maximizing modularity. These statistics are computed on each of the 500 bootstrap samples with replacement of size 70% from dataset, used to compute percentile confidence intervals with error $\alpha = 0.05$ for each statistic. Considering that modularity measures the level of segmentation into modules of the network, and that AIC measures the cohesiveness of these modules, taken together, these plots say that the networks with the most cohesive modules can be found among PF networks.

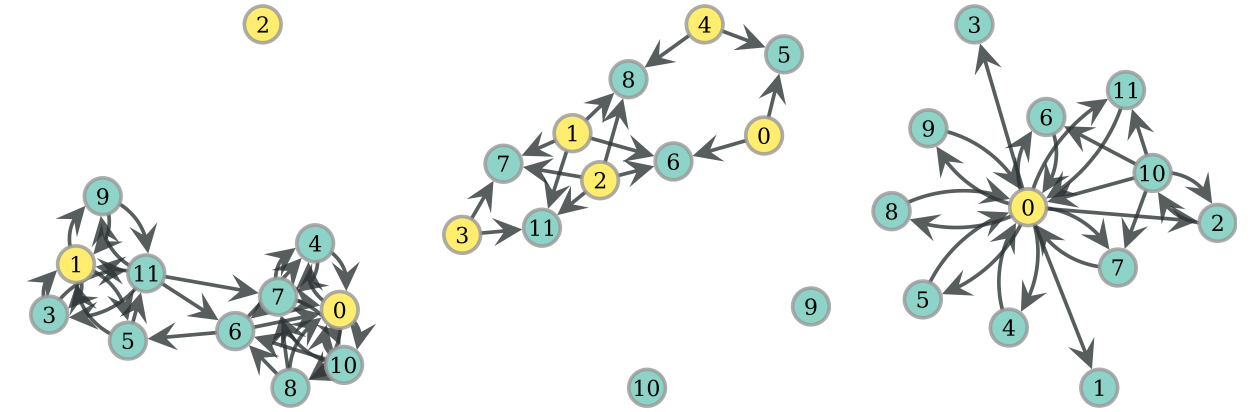
values of 0.083, 0.313 and 0.023, respectively. However, if one looks at the highest values of this measure in Figure 4.10(b), it is noted PF gets higher values than RV, for $ph \leq 0.15$. This pattern is due to networks such as that of Figure 4.12 (a), where there are two modules of connected nodes around distinct hunters, which is reflected in a value of 0.454 for intra-partition clustering. On the other hand, if community structure is measured with modularity index, it can be seen in Figure 4.10(c) that PF mean increases with ph, and it is higher than that of single optima networks. This behavior is influenced mainly by networks that indeed can exhibit relatively large modularity, but with weakly connected modules, similar to the architecture of some RV optima. Figure 4.12 (b) displays an example of this type of PF optima.

Reciprocity on PF optimal networks. The behavior of the other social network pattern we study, reciprocity, can be appreciated in Figure 4.13. There it can be seen that, particularly, the presence of reciprocated arcs between a hunter and non-hunters is very salient on PF networks compared to the other optima. This pattern is observed in networks such as that of Figure 4.12 (c). On the other hand, the presence of reciprocity among non-hunters in Figure 4.13(b), is associated with networks similar to that of Figure 4.11 (b), which for example has 5 reciprocated arcs among non-hunters. The intuition for the presence of reciprocated arcs between a hunter and non-hunters, as well as other patterns found on PF optima, is given by the notion of dominance, which defines the Pareto optimal set, and will be explained next through the example of Figure 4.14.



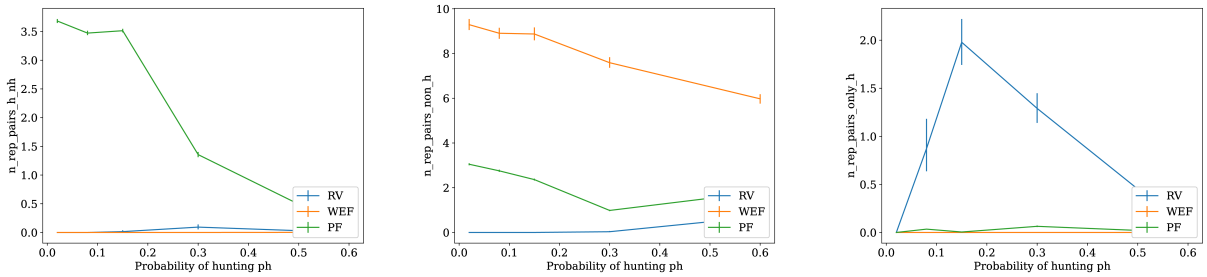
(a) $F = 3$, $nh = 3, ph = 0.02$ (b) $F = 3$, $nh = 3, ph = 0.15$ (c) $F = 4$, $nh = 3, ph = 0.08$

Figure 4.11: **Representative PF networks of regime** $ph \leq 0.15$**.** Model variables under which these Pareto optimal networks were obtained are displayed under each network. Hunters are filled in yellow. To obtain representative networks, these were chosen as the most central from leaves of efficient trees with a relative proportion higher than 8%. To see the trees, consult the Annex D.8. These networks show that, even with tiny chances of hunting (ph) as in (a), PF networks are able to organize into more cooperative schemes, going beyond the hunter homophilic bonds of RV optima (see Section 2.4.4).



(a) $F = 4, nh = 3, ph = 0.08$ (b) $F = 2, nh = 5, ph = 0.6$ (c) $F = 4, nh = 1, ph = 0.15$

Figure 4.12: **Representative PF networks with some social network pattern.** Model variables under which these Pareto optimal networks were obtained are displayed under each network. Hunters are filled in yellow. Average intra-partition clustering (AIC) is the mean of local clustering computed in each subnetwork induced by the modules of the partition maximizing modularity. From left to right: network (a) has a modularity of 0.392 and AIC of 0.454 (highly cohesive modules around hunters), network (b) has a modularity of 0.221 and AIC of 0 (non-cohesive modules), while network (c) has a modularity of 0 and AIC of 0.029 (no modules, and high hunter/non-hunter reciprocity). Additionally, this last network has 7 reciprocated arc pairs between hunters and non-hunters.

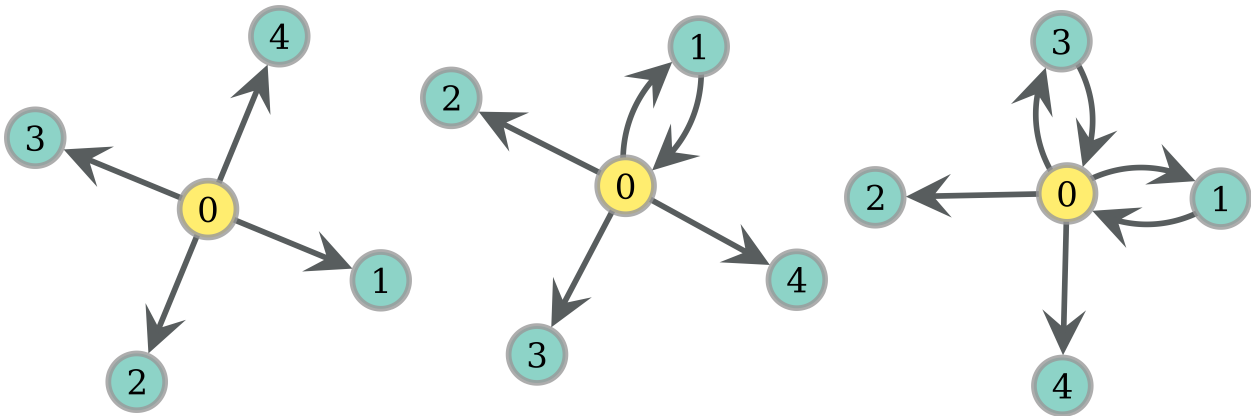


(a) Reciprocated arc pairs between a hunter and a non-hunter.

(b) Reciprocated arc pairs between non-hunters.

(c) Reciprocated arc pairs between hunters.

Figure 4.13: **Distributions of different types of reciprocated arc pairs for each type of network optima.** From left to right: Mean number of reciprocated pairs between a hunter and a non-hunter (left), mean number of reciprocated pairs between non-hunters (center), and mean number of reciprocated pairs between only hunters (right). These means are computed on each of the 500 bootstrap samples with replacement of size 70% from dataset, which are used to compute percentile bootstrap confidence intervals with error $\alpha = 0.05$ for each statistic. The plots show that each type of reciprocated arc, inter-hunter-non-hunter in (a), intra-non-hunter in (b), and intra-hunter in (c), achieves more prevalence in a distinct optima type, respectively in PF, WEF, RV. This is consistent with what is shown in Figure 4.12, Figure 4.4, Figure 4.9, and we interpret it as an impact of the social context on the social organization, see Section 4.4.



(a) Network D_1.

(b) Network D_2.

(c) Network D_3.

Figure 4.14: **Three networks with increasing number of reciprocated arcs between hunter and non-hunters.** We assume there is 1 hunter, node 0, $ph = 0.08$ and $F = 4$. Network D_1 gets costs $RV = 0.261$ and $WEF = 0.024$, D_2 obtains $RV = 0.315$, $WEF = 0.02$ and D_3, $RV = 0.332$, $WEF = 0.018$. The deletion of each additional reciprocated arc cannot simultaneously lower both costs, making networks D_2 and D_3 good candidates to be Pareto optimal.

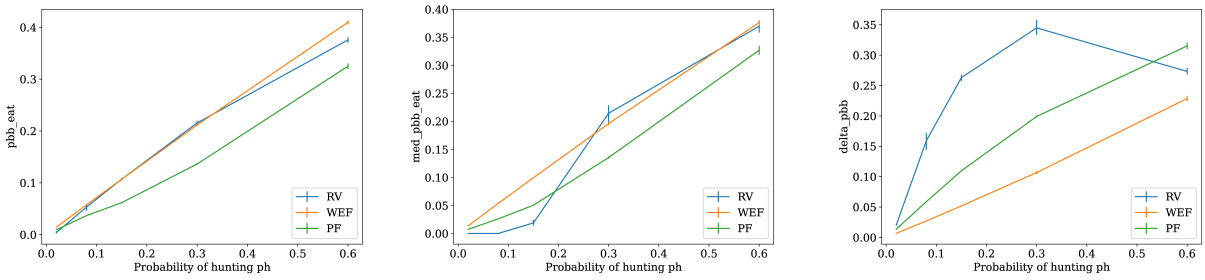
Rationale of traits present on PF networks. Since $ph = 0.08$ for every network in Figure 4.14, in network D_1 each non-hunter eats with equiprobability 0.02. If we add one reciprocated arc pair on node 1, we obtain the network D_2 in Figure 4.14. And since $F = 4$, now each non-hunter in D_2 eats directly from the hunter, and also from the food that first went through node 1. Thus, node 1 eats with probability 0.02 while every other non-hunter eats now with probability 0.035. This reduces the WEF cost, and since we are in this range of low probabilities, the mean of RV costs increases. See the exact values in the Figure description. If now we add a new reciprocated arc pair to node 3 we get the network D_3 in Figure 4.14, where now node 1 eats additionally from the indirect route through 3. This implies that every non-hunter in D_3 eats with a probability of 0.035, which again increases RV and lowers WEF. Thus, for this condition of model variables and these simple networks, the presence of each of these reciprocated arc pairs delivers a non-dominated network, since the deletion of this trait cannot simultaneously lower both costs, as we have seen.

Probability of eating. As a final observation, we look at the behavior of the probability of eating in a network. The mean probability (over network nodes) is on Figure 4.15(a), and it turns out that PF obtains lower means over dataset samples. This is due to the high prevalence of networks with one hunter and reciprocated arcs between hunter and non-hunters, which usually get low mean probabilities, due to low food availability. If now we look at the median probability on Figure 4.15(b), it is observed that PF offers a greater probability than RV on low ph values. This is because in PF there are networks similar to WEF minima that deliver better probabilities of eating that do not leave isolated nodes as with the RV optima. This is confirmed by Figure 4.15(c), which shows the greatest differences between hunters and non-hunters are in the RV optima, and by the remarkable fact in Figure 4.15(b) that WEF optima shows the greatest efficiency in the probability of eating in the whole range of ph.

4.4 Discussion

In this section, we discuss the main contributions of our model in relation to previous works, possible implications of our results, and avenues for future research.

Condition for community network structure. Communities given by cohesive network modules appear in food sharing networks optimized to provide egalitarian access to food resources jointly with reduced individual risk. This result is obtained under conditions of uncertainty and scarcity in the food supply (see Figure 4.12, Figure 4.10(b)), which suggests their adaptive value for this context, and can be considered a new mechanism for the emergence of communities in networks, as those reviewed in Section 2.4 from the survey chapter. The simultaneous optimization of the two goals is necessary since RV optima may present community structure but with rather sparse connectivity (see Figure 4.9), while WEF may present communities under the special case of degree dispersion minimization (see Figure 4.5). These PF networks are resemblant to the food-sharing networks observed by Dyble et al (2016) [89], where there are cohesive groups corresponding to households pro-



(a) $\bar{pe}$. (b) $med(pe)$. (c) $\bar{pe}_h - \bar{pe}_{nh}$.

Figure 4.15: **Display of three variables associated with the probability of eating for each type of network optima.** From left to right: mean probability of eating, median probability of eating, and difference in means of the probability of eating between hunters and non-hunters. In each graph, the mean of each statistic is drawn surrounded by percentile bootstrap confidence intervals with error $\alpha = 0.05$, computed on each of the 500 bootstrap samples with replacement of size 70% from the dataset. Overall, PF networks for low ph values are more resilient than RV, since may organize into more egalitarian (WEF-like) structures. WEF networks deliver the greatest efficiency in most of the range of ph, either in terms of the median pe, or the differences in $\bar{pe}$ between hunters and non-hutners.

visioned by an adult couple, and a set of households forms a cluster with a small number of inter-household connections. In our model, the cohesive groups correspond to consumers, or non-hunters, grouped around one or more hunters. This result suggests that not only the camp level, which groups several clusters, is functionally associated with risk reduction reciprocity as argued by Dyble et al (2016), but also the hierarchically lower levels of households and clusters may have been related evolutionarily to some resource distribution optimality that includes the egalitarian access to resources. This idea is consistent with some studies suggesting that many kinship configurations may be a product of sharing patterns [180, 136, 152].

The claim that the two minimization criteria are required for the emergence of cohesive communities is reinforced by the condition of uncertain food generation, not only because it is under this regime where communities are obtained (Figure 4.12), but also because when food supply is safer, the two criteria become indistinct to the dense optimal WEF networks with low modularity (Figure 4.9(f)). The fact that sharing is directed mainly towards large and uncertain packages of food is acknowledged by several empirical studies [122, 145, 34, 130]. The necessity of the confluence of both the individual and group interests has also been pointed out in the simulation agent-based study by Briz i Godino et al (2014) [50] which models the cooperative public calls through smoke signals of a hunter-fisher-gatherer society conditional on the exceptional accumulation of fish. The model shows that a low weight on social capital relative to individual consumption of meat in the agents' fitness, is enough to obtain most society cooperating as the most likely outcome, but this result may be reversed when this weight is absent. This social capital component of reproductive fitness seems to be a plausible mechanism according to Chaudhary et al (2016) [63], where it is reported that relational wealth, or the cooperative relationships an individual has, provides advantages in

buffering food risk, body mass index, fertility, and is partially heritable.

Finally, the result that a stringent resource availability would be necessary for the evolution of cooperation is supported by another agent-based model named *Cooperation under resource pressure* (CURP) by Pereda et al (2017) [227]. This work shows that under moderate survival stress, populations of agents self-organize in an indirect reciprocity system consisting in sharing the part of the resource that is not strictly necessary for survival, achieving to collectively lowering the chances of starving. Additionally, a high-stress regime turns into unstable behavior, where the population constantly searches for survival strategies, while low-stress does not exert selection and strategies remain almost constant and randomly drifting. Since the variables in our model may be understood as a form of resource-pressure, and even a parallel may be established between the two models' variables (*prob-resource* to ph, and *min-energy* to F), the results of our model roughly resemble those of CURP, taking into account that CURP is not a social networks model, and that ours is not an agent-based evolutionary model. This aclaration allows us to understand the fact that our model does not produce a clear distinction between moderate and high-stress scenarios in the sense CURP does. Our model does not incorporate a notion of dynamics like evolutionary models, but rather gives a fixed optimal organization under certain conditions. On the other hand, CURP does not make a distinction between hunters and non-hunters, and all agents can produce food.

Network structures promoting egalitarian resource distribution. Welfare optimal networks (WEF) display high connectivity but low network segmentation, where each hunter is connected to one big, homogeneous and dense group of non-hunters, see Figure 4.4. On the other hand, networks minimizing individual risk (RV) range from cooperation exclusive among hunters, to networks more similar to welfare networks when the food supply is safer, see Figure 4.9. These networks are very similar to those tested in the lab experiment by Chiang (2015) [64], where it is found that network structures linking agents with discrepant income levels promote more egalitarian distributions by motivating the rich agents to share their incomes with the poor. Particularly, the network *SF_negative* from the experiment is very similar to the undirected version of WEF optima from our model. On the contrary, networks where similar income agents are linked evolve to more unequal income distributions, which is also consistent with the RV optima in stringent resource conditions where hunters are only connected among themselves. This result is consistent with the view that social networks structured originally according to ecological considerations, may have created the environment in which prosocial tendencies and equity response elicitation as those explored in the experiment, evolved along human history [144, 139].

Finally, the fact that many WEF optima minimize degree variance is consistent with results of a previous simulation of need-based transfers evolution model showing that degree variance is anti-correlated with survival rate [160]. The minimization of in-degree variance is another pathway to cohesive communities found in our model, see Fig. 4.5. This scenario may be the idealized case resultant of some other more realistic constraint as, for example, an upper bound on the number of connections an agent may have. This last idea has been proposed by the *social brain hypothesis* [191], which states that there are cognitive constraints on the number of face-to-face social interactions a human may have, where these limits would

be inversely related to the emotional closeness of the bond. Thus, for example, the number of close friends one could maintain is limited to five. The evolutionary perspective suggests, again, that these cognitive constraints on social relationships may have been related originally to some kind of resource distribution optimality.

Distinct patterns of reciprocity. We have obtained distinct distributions of reciprocity among hunters and non-hunters, in the three types of optimal networks we study, which gives a broader picture to the usual notion that reciprocity is driven by the minimization of food production uncertainty [155, 290]. We only got high reciprocity under the reduction of variability criterion (RV) for reciprocal exchanges among hunters, see Figure 4.13(c). On the other hand, Pareto optimal networks (PF) maximize mean reciprocity between hunters and non-hunters, and Welfare (WEF) maximizes mean reciprocity among non-hunters, see Figure 4.13. These distinct distributions may suggest that the relative importance of every criterion guiding food transfers depends critically upon the context, which in turn may be a result of population size and division of labor.

One possibility is that for large communities, where hunters are only a small fraction of the whole population, there are hunters that reciprocally exchange food to minimize risk, and after that, new exchanges take place within subcommunities fed by each of these hunters, that may be driven to a greater or lesser extent by welfare considerations. This is not far from what has been reported for some hunter-gatherer societies, where there are two systems or stages of sharing: in the first, sharing occurs among participants in the cooperative effort of food acquisition as a form of labor reward, and then, in the secondary distribution each individual that received shares redistributes his or her share to families that did not participate [153]. This is consistent with the claim that exchanges motivated by need or welfare, would be more common within a household [169, 127], and with more recent observations reporting that Hadza men consumed a substantial amount of food while out of camp foraging [30], and that returning to camp empty-handed is indicative that he failed to produce enough surplus to share. This may suggest a greater relative weight of individual starvation risk minimization over welfare motivation in the first stage of sharing.

Implications for evolutionary and economic models of social dynamics, and future work. Though our model is not evolutionary in its formulation, its results suggest that motives not usually modeled by these approaches, such as the egalitarian group access to critical resources for survival, may be an important driver for social network formation. Typical evolutionary models of food-sharing [151] rely on the assumption of maximization of a function of subjective preferences, usual in economic models of network formation [156], which highlights an exclusive role of individual choice and may impose high cognitive requirements. We claim that evolutionary approaches may benefit from a wider repertoire of assumptions including a resource distribution perspective, the modeling of survival needs, and the explicit inclusion of the group level of analysis.

Our model looks for capturing the transfers of finite resources subject to uncertain and scarce production, and their effect on optimal network organization. Therefore, its insights may potentially be applied to other situations imposing an analog structure of restraints. For example, since parents need to spend most of their time raising their offspring, maybe

the behavior of cooperative breeding is dependent on the uncertain generation of scarce free time available for the care of non-descendant children. On the other hand, some works have argued [257, 120] about the usefulness of using a diminishing marginal returns function over the exchanged quantities of food, which may be a symptom of the underlying phenomenon of satisfaction of needs [106]. Our model is based on an original formulation of starvation risk relying upon the expected number of success runs of a certain length in a sequence of trials to implement the hunger need over time. Our model may contribute a formal framework to proceed in this discussion. Finally, the explicit inclusion of group level of analysis may be implemented through the multi-level selection model [289, 215] that incorporates as evolutionary units the individuals, as well as the groups they form. A mapping from individual fitness to starvation risk, and from group fitness to egalitarian group access to resources may be studied in future work. A model like this may serve to shed light on the complex interactions between developmental, biological, social, economic and cultural factors influencing social networks formation. More direct extensions of our model would be, for example, more complex protocols of food-sharing allowing for preferential exchanges according to agent closeness, correlations in production of food [290], or others. The modular character of the model would allow making changes at several levels.

4.5 Conclusion

We have presented and analyzed a formal network optimization model inspired by a food-sharing dynamic, that can recover some typical patterns found in social networks. Specifically, we have formalized two main drivers for food-sharing: the reduction of individual starvation risk and the care for the general welfare of agents, and have shown using evolutionary algorithms and data analysis techniques, three main findings, plus a methodological contribution.

- Communities of cohesive network modules appear in food sharing networks optimized to provide egalitarian access jointly with reduced individual risk. This result is obtained under conditions of uncertainty and scarcity in the food supply, which suggests their adaptive value for this context. The simultaneous optimization of the two goals is necessary for this outcome. These networks are resemblant to the food-sharing networks observed by Dyble et al (2016) [89], where there are cohesive groups corresponding to households provisioned by an adult couple, and a set of households forms a cluster with a small number of inter-household connections. In our model, the cohesive groups correspond to consumers, or non-hunters, grouped around one or more hunters. This result suggests that the organization of nuclear families described by Dyble et al (2016) [89] may have been evolutionarily functional for resource distribution optimality.
- Welfare optimal networks (WEF) display high connectivity, but low network segmentation, where each hunter is connected to one big, homogeneous and dense group of non-hunters. On the other hand, networks minimizing the individual risk (RV), which range from cooperation exclusive among hunters to networks more similar to welfare networks when food supply is safer, may present community structure but with rather sparse connectivity. These networks are very similar to those tested in the lab experiment by Chiang (2015) [64], where it is found that network structures linking agents

with discrepant income levels promote more egalitarian distributions by motivating the rich agents to share their incomes with the poor. Particularly, network *SF_negative* from the experiment is very similar to the undirected version of WEF optima from our model. On the contrary, networks where similar income agents are linked evolve to more unequal income distributions, which is also consistent with RV optima in stringent resource conditions where hunters are only connected among themselves. This result is consistent with the view that social networks structured originally according to ecological considerations, may have created the environment in which prosocial tendencies and equity response elicitation as those explored in the experiment, evolved along human history [144, 139].

- We have obtained distinct distributions of reciprocity among hunters and non-hunters, in the three types of optimal networks we study, which gives a broader picture of the usual notion that reciprocity is driven by the minimization of food production uncertainty [155], and that may be consistent with some empirical reports [153] on how sharing is distributed in waves, first among hunters, and then hunters with their families.

- As a final contribution regarding methodology, our model is based on an original formulation of survival risk relying upon an estimate of the number of success runs of a certain length, in a sequence of Bernoulli trials. We believe our model may contribute a formal framework to proceed in this discussion regarding the principles guiding food-sharing network formation. Additionally, we employ an original pipeline of state-of-art clustering algorithms to analyze the multiple network optima of our model that may be of interest.

Our model suggests that evolutionary accounts of food sharing may benefit from including a resource distribution perspective, the modeling of survival needs, and the explicit inclusion of the group level of analysis, for example, as a level where selection also operates, autonomously from the individual level. Future work based on this model may contribute to a better understanding of the complex interaction of factors affecting the formation of social networks, and human natural history.

4.6 Acknowledgements

The authors would like to thank Renato Cerro, who is professional editor in the Department of Computer Science of Universidad de Chile, by his work of copyediting of the manuscript in the review stage of the paper displayed in this chapter.

Chapter 5

Conclusion

We end this document with a review of the distinct contributions and takeaway points of the presented works.

QuickCent results. The results presented in Chapter 3 show that QuickCent can be a competent alternative to make a regression on a power-law centrality variable. The method generates **accurate** and **low variance** estimates even if trained with a **small training set**, comparable in precision to some more advanced machine learning algorithms. Its accuracy is available at a **lower time cost** than more complex machine learning methods. In this sense, QuickCent is an example of simple heuristics based on the **ecological validity**, or the exploitation of regularities of the data that makes the application of the heuristics be sound or incorrect, that can be a competitive alternative to more computationally intensive procedures. This ecological validity is a way to call the cost to pay, in terms of the bias-variance tradeoff, when using rigid low-variance heuristics.

Approximation of size measures on some scale-free networks. The study case addressed in the paper corresponds to the approximation of expensive harmonic centrality by cheap in-degree centrality, and more generally, in terms of the axioms of centrality by Boldi and Vigna (2014) [40], the **approximation** of an expensive centrality sensitive to size and density **by a cheap density measure**. The comparison between the performance on synthetic networks produced with preferential attachment and some empirical networks, shows that, since **preferential attachment** can be produced by link creation processes guided by the **local network structure** such as meeting friends of friends, the local density could indeed **reflect greater group size**. We think that this insight may be of utility for other cases, such as estimating other centrality measures, or to find footprints of the network generation mechanism.

Food-sharing modeling results. We have found that communities of **cohesive network modules** appear in food sharing networks optimized to provide **egalitarian access jointly with reduced individual risk**. This result is obtained under conditions of uncertainty

and scarcity in the food supply, which suggests their adaptive value for this context. These networks are resemblant to the food-sharing networks observed by Dyble et al (2016) [89], where there are cohesive groups corresponding to households provisioned by an adult couple, and a set of households forms a cluster with a small number of inter-household connections. Now, **welfare optimal networks** (WEF) display **high connectivity, but low network segmentation**, where each hunter is connected to one big, homogeneous and dense group of non-hunters. On the other hand, **networks minimizing the individual starvation risk** (RV), range from **cooperation exclusive among hunters** to networks more similar to welfare networks when food supply is safer. These networks are very similar to those tested in the lab experiment by Chiang (2015) [64], where it is found that network structures linking agents with discrepant income levels promote more egalitarian distributions by motivating the rich agents to share their incomes with the poor. On the contrary, networks where similar income agents are linked evolve to more unequal income distributions, which is also consistent with RV optima in stringent resource conditions where hunters are only connected among themselves. Finally, we have obtained **distinct distributions of reciprocity among hunters and non-hunters**, in the three types of optimal networks we study, which gives a broader picture of the usual notion that reciprocity is driven by the minimization of food production uncertainty [155], and that may be consistent with some empirical reports [153] on how **sharing is distributed in waves**, first among hunters, and then hunters with their families.

Interpretation of food-sharing results. All the modeling results are consistent to the argued relevance that food sharing would have for the **evolution** of different forms of **human cooperation**. **Social networks** structured originally according to ecological considerations, may have created the **environment** in which, **prosocial tendencies** and equity response elicitation, may be potentially **elicited and evolved** along human history. On the other hand, the model results have implications regarding the usual approaches adopted by evolutionary models. We claim that **evolutionary approaches** may benefit from a **wider repertoire of assumptions** including a **resource distribution** perspective, the modeling of **survival needs**, and the explicit inclusion of the **group level of analysis**.

Food-sharing methodological contributions. Our model is based on an **original formulation of starvation risk** relying upon an estimate of the number of success runs of a certain length, in a sequence of Bernoulli trials. We believe our model may contribute a formal framework to proceed in this discussion regarding the principles guiding food-sharing network formation. Additionally, we employ an **original pipeline of state-of-art clustering algorithms**, that is, the application of tSNE algorithm [275] followed by OPTICS [10], to analyze the multiple network optima of our model that may be of interest.

Chapter 6

Declarations

6.1 Funding.

This work was supported in part by the National Center for Artificial Intelligence CENIA [FB210017, Basal ANID], Fondecyt grant 1200967, the Millennium Institute Foundational Research on Data IMFD, the supercomputing infrastructure of the NLHPC, National Laboratory for High Performance Computing (ECM-02), the National Agency of R+D ANID doctoral studies scholarship [CONICYT-PCHA/Doctorado Nacional/2016-21161085], and the doctoral studies scholarship by the Escuela de Postgrado y Educación Continua, Facultad de Ciencias Físicas y Matemáticas Universidad de Chile. The funding sources played no role in the research.

6.2 Competing interests.

The authors have no competing interests to declare that are relevant to the content of this thesis.

6.3 Data availability Statement.

All datasets and code used in this work are publicly available.

For Chapter 3 and Annex B, they are available under the following DOI and private link (for review purposes). If the paper is accepted, the repository [231] produces a new public link, and the DOI is published.

`10.6084/m9.figshare.22183372`
`https://figshare.com/s/64b43eb2d75e64e72ca1`

For Chapter 4 and Annex D, all the code written and datasets generated and used in the analysis of this work, are available at the following DOI and repository [233].

`10.6084/m9.figshare.19203926`
`https://figshare.com/articles/software/food_sharing_data_and_code/19203926`

For Annex E, all the code, datasets, as well as the documentation for running the methods to get the results presented, are available in the following link.

`https://github.com/PanchoTonho/Scale_free_food_sharing`

6.4 Author contributions.

For Chapter 3, all authors contributed to the study conceptualization, formal analysis, methodology, resources, funding acquisition, validation, visualization and writing. F.P. additionally performed data curation, investigation and software implementation. J.P. and A.A. contributed supervision and project administration.

For Chapter 4, all authors contributed to the study conceptualization, funding acquisition, project administration and writing. F.P. and J.P. contributed to the formal analysis, methods, validation and visualization. F.P. additionally performed data curation, investigation, and software implementation. J.P. and A.A. contributed supervision.

Bibliography

[1] Alireza Abbasi, Liaquat Hossain, and Loet Leydesdorff. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6(3):403–412, 2012.

[2] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

[3] V. Ahedo, J. Caro, E. Bortolini, D. Zurro, M. Madella, and J. M. Galán. Quantifying the relationship between food sharing practices and socio-ecological variables in small-scale societies: A cross-cultural multi-methodological approach. *PloS one*, 14(5):e0216302, 2019.

[4] Réka Albert. Scale-free networks in cell biology. *Journal of cell science*, 118(21):4947–4957, 2005.

[5] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *nature*, 401(6749):130–131, 1999.

[6] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *nature*, 401(6749):130–131, 1999.

[7] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

[8] Wesley Allen-Arave, Michael Gurven, and Kim Hill. Reciprocal altruism, rather than kin selection, maintains nepotistic food transfers on an ache reservation. *Evolution and Human Behavior*, 29(5):305–318, 2008.

[9] Attila Ambrus, Markus Mobius, and Adam Szeidl. Consumption risk-sharing in social networks. *American Economic Review*, 104(1):149–82, 2014.

[10] Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: Ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2):49–60, 1999.

[11] Coren L Apicella, Frank W Marlowe, James H Fowler, and Nicholas A Christakis. Social networks and cooperation in hunter-gatherers. *Nature*, 481(7382):497–501, 2012.

[12] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.

[13] Lars G Backlund, Johan Bring, Ylva Skånér, Lars-Erik Strender, and Henry Montgomery. Improving fast and frugal modeling in relation to regression analysis: Test of 3 models for medical decision making. *Medical decision making*, 29(1):140–148, 2009.

[14] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42. ACM, 2012.

[15] Robert Converse Bailey. *The behavioral ecology of Efe Pygmy men in the Ituri Forest, Zaire*, volume 86. University of Michigan Museum, 1991.

[16] P. Bak. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus, New York, 1996.

[17] Felipe Balmaceda and Juan F Escobar. Trust in cohesive communities. *Journal of Economic Theory*, 170:289–318, 2017.

[18] Albert-László Barabási. *Network science*. Cambridge University Press, 2016.

[19] Albert-László Barabási. Love is all you need: Clauset's fruitless search for scale-free networks. *Blog post available at https://www. barabasilab. com/post/love-is-all-you-need*, page 20, 2018.

[20] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[21] Albert-Laszlo Barabâsi, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4):590–614, 2002.

[22] A.L. Barabási. *Network science*. Cambridge Univ. Press, 2016.

[23] Alain Barrat and Martin Weigt. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560, 2000.

[24] Marc Barthélémy and Luis A Nunes Amaral. Small-world networks: Evidence for a crossover picture. *Physical Review Letters*, 82(15):3180, 1999.

[25] Heiko Bauke, Cristopher Moore, Jean-Baptiste Rouquier, and David Sherrington. Topological phase transition in a network model with preferential attachment and node removal. *The European Physical Journal B-Condensed Matter and Complex Systems*, 83(4):519–524, 2011.

[26] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini. Emergence of polarized ideological opinions in multidimensional topic spaces. *Physical Review X*, 11(1):011012, 2021.

[27] Patrick Bayer, Stephen L Ross, and Giorgio Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of political Economy*, 116(6):1150–1196, 2008.

[28] Anna C Belkina, Christopher O Ciccolella, Rina Anno, Richard Halpert, Josef Spidlen, and Jennifer E Snyder-Cappione. Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature communications*, 10(1):1–12, 2019.

[29] Kamal Berahmand, Asgarali Bouyer, and Negin Samadi. A new local and multidimensional ranking measure to detect spreaders in social networks. *Computing*, 101:1711–1733, 2019.

[30] J Colette Berbesque, Brian M Wood, Alyssa N Crittenden, Audax Mabulla, and Frank W Marlowe. Eat first, share later: Hadza hunter–gatherer men consume more while foraging than in central places. *Evolution and Human Behavior*, 37(4):281–286, 2016.

[31] Noam Berger, Christian Borgs, J Chayes, R D'Souza, and R Kleinberg. Competition-induced preferential attachment. *Automata, languages and programming*, pages 208–221, 2004.

[32] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[33] T. C. Bergstrom. The algebra of assortative encounters and the evolution of cooperation. *International Game Theory Review*, 5(03):211–228, 2003.

[34] R. B. Bird, D. W. Bird, E. A. Smith, and G. C. Kushnick. Risk and reciprocity in meriam food sharing. *Evolution and Human Behavior*, 23(4):297–321, 2002.

[35] C.M. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006.

[36] Francis Bloch, Garance Genicot, and Debraj Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143(1):36–58, 2008.

[37] Marián Boguná, Fragkiskos Papadopoulos, and Dmitri Krioukov. Sustaining the internet with hyperbolic mapping. *Nature communications*, 1(1):1–8, 2010.

[38] Marián Boguná and Romualdo Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112, 2003.

[39] Marián Boguná, Romualdo Pastor-Satorras, and Alessandro Vespignani. Cut-offs and finite size effects in scale-free networks. *The European Physical Journal B*, 38(2):205–209, 2004.

[40] Paolo Boldi and Sebastiano Vigna. Axioms for centrality. *Internet Mathematics*, 10(3-4):222–262, 2014.

[41] Béla Bollobás. Degree sequences of random graphs. *Discrete Mathematics*, 33(1):1–19, 1981.

[42] Béla Bollobás* and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[43] Béla Bollobás, Oliver Riordan, Joel Spencer, Gábor Tusnády, et al. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[44] Béla Bollobás and Oliver M Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pages 1–34, 2003.

[45] Vincent Boucher. Structural homophily. *International Economic Review*, 56(1):235–264, 2015.

[46] Yann Bramoullé, Rachel Kranton, and Martin D'amours. Strategic interaction and networks. *American Economic Review*, 104(3):898–930, 2014.

[47] Ulrik Brandes and Christian Pich. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(07):2303–2318, 2007.

[48] Leo Breiman, JH Friedman, Richard A Olshen, and Charles J Stone. Classification and regression trees. wadsworth, 1984. *Intelligence*, pages 1002–1007, 1993.

[49] Henry Brighton and Gerd Gigerenzer. The bias bias. *Journal of Business Research*, 68(8):1772 – 1784, 2015. Special Issue on Simple Versus Complex Forecasting.

[50] Ivan Briz i Godino, José Ignacio Santos, José Manuel Galán, Jorge Caro, Myrian Álvarez, and Débora Zurro. Social cooperation and resource management dynamics among late hunter-fisher-gatherer societies in tierra del fuego (south america). *Journal of Archaeological Method and Theory*, 21(2):343–363, 2014.

[51] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1):309–320, 2000.

[52] Arndt Bröder and Ben Newell. Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, 3(3):205, 2008.

[53] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

[54] Guido Caldarelli. *Scale-Free Networks: Complex Webs in Nature and Technology (Oxford Finance)*. Oxford University Press, USA, 2007.

[55] Guido Caldarelli, Andrea Capocci, Paolo De Los Rios, and Miguel A Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, 89(25):258702, 2002.

[56] Ettore Camerlenghi, Alexandra McQueen, Kaspar Delhey, Carly N Cook, Sjouke A Kingma, Damien R Farine, and Anne Peters. Cooperative breeding and the emergence of multilevel societies in birds. *Ecology letters*, 25(4):766–777, 2022.

[57] Ramon Cancho and Ricard Solé. Optimization in complex networks. *Statistical mechanics of complex networks*, pages 114–126, 2003.

[58] Jean M Carlson and John Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2):1412, 1999.

[59] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[60] Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011.

[61] Damon Centola, Juan Carlos Gonzalez-Avella, Victor M Eguiluz, and Maxi San Miguel. Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905–929, 2007.

[62] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, 38(1):2, 2006.

[63] Nikhil Chaudhary, Gul Deniz Salali, James Thompson, Aude Rey, Pascale Gerbault, Edward Geoffrey Jedediah Stevenson, Mark Dyble, Abigail E Page, Daniel Smith, Ruth Mace, et al. Competition for cooperation: variability, benefits and heritability of relational wealth in hunter-gatherers. *Scientific Reports*, 6(1):1–7, 2016.

[64] Y. S. Chiang. Good samaritans in networks: An experiment on how networks influence egalitarian sharing and the evolution of inequality. *PloS one*, 10(6):e0128777, 2015.

[65] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[66] Michael Suk-Young Chwe. Structure and strategy in collective action. *American journal of sociology*, 105(1):128–156, 1999.

[67] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[68] Aaron Clauset, Maxwell Young, and Kristian Skrede Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58–87, 2007.

[69] Carlos A Coello Coello, Gary B Lamont, David A Van Veldhuizen, et al. *Evolutionary algorithms for solving multi-objective problems*, volume 5. Springer, 2007.

[70] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Physical review letters*, 85(21):4626, 2000.

[71] James S Coleman. Social capital in the creation of human capital. *American journal of sociology*, 94:S95–S120, 1988.

[72] Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical review letters*, 99(14):148701, 2007.

[73] Ewan R Colman and Geoff J Rodgers. Complex scale-free networks with tunable power-law exponent and clustering. *Physica A: Statistical Mechanics and its Applications*, 392(21):5501–5510, 2013.

[74] Pol Colomer-de Simón, M Serrano, Mariano G Beiró, J Ignacio Alvarez-Hamelin, and Marián Boguná. Deciphering the global organization of clustering in real complex networks. *Scientific reports*, 3(1):1–7, 2013.

[75] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, Cambridge, Massachusetts, 2 edition, 2001.

[76] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[77] Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

[78] Ithiel de Sola Pool and Manfred Kochen. Contacts and influence. *Social networks*, 1(1):5–51, 1978.

[79] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[80] Maxime Derex and Robert Boyd. Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, 113(11):2982–2987, 2016.

[81] Maxime Derex, Bernard Godelle, and Michel Raymond. Social learners require process information to outperform individual learners. *Evolution: International Journal of Organic Evolution*, 67(3):688–697, 2013.

[82] Mandeep K Dhami and Clare Harries. Fast and frugal versus regression models of human judgement. *Thinking & Reasoning*, 7(1):5–27, 2001.

[83] Sergei N Dorogovtsev, José FF Mendes, and Sergei N Dorogovtsev. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2003.

[84] Sergey N Dorogovtsev and José Fernando F Mendes. Evolution of networks with aging of sites. *Physical Review E*, 62(2):1842, 2000.

[85] Sergey N Dorogovtsev, José Fernando F Mendes, and Alexander N Samukhin. Structure of growing networks with preferential linking. *Physical review letters*, 85(21):4633, 2000.

[86] Jonathan PK Doye. Network topology of a potential energy landscape: A static scale-free network. *Physical review letters*, 88(23):238701, 2002.

[87] Holger Drees, Anja Janßen, Sidney I Resnick, and Tiandong Wang. On a minimum distance procedure for threshold selection in tail analysis. *SIAM Journal on Mathematics of Data Science*, 2(1):75–102, 2020.

[88] Kevin Dunne, Padraig Cunningham, and Francisco Azuaje. Solutions to instability problems with sequential wrapper-based approaches to feature selection. *Journal of Machine Learning Research*, 1:1–22, 2002.

[89] Mark Dyble, James Thompson, Daniel Smith, Gul Deniz Salali, Nikhil Chaudhary, Abigail E Page, Lucio Vinicuis, Ruth Mace, and Andrea Bamberg Migliano. Networks of food sharing reveal the functional significance of multilevel sociality in two hunter-gatherer groups. *Current Biology*, 26(15):2017–2021, 2016.

[90] Bradley Efron and Robert Tibshirani. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

[91] Agoston E Eiben and James E Smith. *Introduction to evolutionary computing*. Springer, 2015.

[92] Paul Erdös and Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[93] Paul Erdös and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[94] Paul Erdös and Alfréd Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 12(1-2):261–267, 1961.

[95] Alex Fabrikant, Elias Koutsoupias, and Christos H Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *International Colloquium on Automata, Languages, and Programming*, pages 110–122. Springer, 2002.

[96] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262, 1999.

[97] William Feller. *An introduction to probability theory and its applications: volume I*, volume 3. John Wiley & Sons New York, 1968.

[98] Andreas Flache and Michael W Macy. Small worlds and cultural polarization. In *Micro-Macro Links and Microfoundations in Sociology*, pages 152–182. Routledge, 2014.

[99] Peter J Fleming. Computer aided control system design using a multi-objective optimization approach. In *Proceedings of Control 1985 Conference, Cambridge, UK*, pages 174–179, 1985.

[100] Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13:2171–2175, jul 2012.

[101] Félix-Antoine Fortin, Simon Grenier, and Marc Parizeau. Generalizing the improved run-time complexity algorithm for non-dominated sorting. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 615–622, 2013.

[102] Félix-Antoine Fortin and Marc Parizeau. Revisiting the nsga-ii crowding-distance computation. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 623–630, 2013.

[103] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[104] Caxton C Foster, Anatol Rapoport, and Carol J Orwant. A study of a large sociogram ii. elimination of free parameters. *Behavioral science*, 8(1):56–65, 1963.

[105] JC Fu and MV Koutras. Distribution theory of runs: a markov chain approach. *Journal of the American Statistical Association*, 89(427):1050–1058, 1994.

[106] S. Fuller. Positivism, history of. In N. J. Smelser and P. B. Baltes, editors, *International encyclopedia of the social and behavioral sciences*, pages 11821–11827. Elsevier, Amsterdam, 2001.

[107] Martin Gerlach and Eduardo G Altmann. Testing statistical laws in complex systems. *Physical Review Letters*, 122(16):168301, 2019.

[108] Robert Gibrat. *Les inégalités économiques*. Recueil Sirey, 1931.

[109] G. Gigerenzer, P.M. Todd, and ABC Research Group. *Simple heuristics that make us smart*. Oxford UP, New York, 1999.

[110] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[111] Colin S. Gillespie. Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software*, 64(2):1–16, 2015.

[112] Edward L Glaeser, David I Laibson, Jose A Scheinkman, and Christine L Soutter. Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

[113] James P Gleeson. Bond percolation on a class of clustered random networks. *Physical Review E*, 80(3):036107, 2009.

[114] Michel L Goldstein, Steven A Morris, and Gary G Yen. Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(2):255–258, 2004.

[115] Michael Golosovsky. Power-law citation distributions are not scale-free. *Physical Review E*, 96(3):032306, 2017.

[116] Benjamin Golub and Matthew O Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012.

[117] Mark Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[118] Robert M. Gray and David L. Neuhoff. Quantization. *IEEE transactions on information theory*, 44(6):2325–2383, 1998.

[119] Enrico Gregori, Luciano Lenzini, and Chiara Orsini. k-dense communities in the internet as-level topology graph. *Computer Networks*, 57(1):213–227, 2013.

[120] M. Gurven. Reciprocal altruism and food sharing decisions among hiwi and ache hunter–gatherers. *Behavioral Ecology and Sociobiology*, 56(4):366–380, 2004.

[121] M. Gurven and K. Hill. Why do men hunt? a reevaluation of "man the hunter" and the sexual division of labor. *Current Anthropology*, 50(1):51–74, 2009.

[122] Michael Gurven. Reciprocal altruism and food sharing decisions among hiwi and ache hunter–gatherers. *Behavioral Ecology and Sociobiology*, 56(4):366–380, 2004.

[123] Michael Gurven, Kim Hill, Hillard Kaplan, Ana Hurtado, and Richard Lyles. Food transfers among hiwi foragers of venezuela: tests of reciprocity. *Human Ecology*, 28(2):171–218, 2000.

[124] Michael Gurven and Adrian V Jaeggi. Food sharing. In R.A. Scott, R.H. Scott, S.M. Kosslyn, and M.C. Buchmann, editors, *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, Wiley Online Library: Books, pages 1–12. John Wiley & Sons, Incorporated, 2015.

[125] Isabelle Guyon, Kristin Bennett, Gavin Cawley, Hugo Jair Escalante, Sergio Escalera, Tin Kam Ho, Núria Macià, Bisakha Ray, Mehreen Saeed, Alexander Statnikov, and Evelyne Viegas. Design of the 2015 chalearn automl challenge. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2015.

[126] Matthew Haag and Roger Lagunoff. Social norms, local interaction, and neighborhood planning. *International Economic Review*, 47(1):265–296, 2006.

[127] Raymond Hames. Reciprocal altruism in yanomamö food exchange. In Lee Cronk, Napoleon A. Chagnon, and William Irons, editors, *Adaptation and human behavior: an anthropological perspective*, pages 397–416. Aldine de Gruyter, Hawthorne, NY, 2000.

[128] Raymond Hames and Carl McCabe. Meal sharing among the ye'kwana. *Human Nature*, 18(1):1–21, 2007.

[129] Marcus J Hamilton. Collective computation, information flow, and the emergence of hunter-gatherer small-worlds. *Journal of Social Computing*, 3(1):18–37, 2022.

[130] K. Hawkes. Showing off: tests of an hypothesis about men's foraging goals. *Ethology and sociobiology*, 12(1):29–54, 1991.

[131] Kristen Hawkes and Rebecca Bliege Bird. Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 11(2):58–67, 2002.

[132] R Hertwig, U Hoffrage, and L Martignon. Quick estimation: Letting the environment do some of the work. *Simple Heuristics that Make Us Smart*, pages 209–234, 1999.

[133] Kim Hill and A Magdalena Hurtado. Cooperative breeding in south american hunter–gatherers. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674):3863–3870, 2009.

[134] Robin M Hogarth and Natalia Karelaia. Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology*, 49(2):115–124, 2005.

[135] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Physical review E*, 65(2):026107, 2002.

[136] Paul L Hooper, Michael Gurven, Jeffrey Winking, and Hillard S Kaplan. Inclusive fitness and differential productivity across the life course determine intergenerational transfers in a small-scale human society. *Proceedings of the Royal Society B: Biological Sciences*, 282(1803):20142808, 2015.

[137] Kurt Hornik, Christian Buchta, and Achim Zeileis. Open-source machine learning: R meets Weka. *Computational Statistics*, 24(2):225–232, 2009.

[138] Reza Hosseini. Quantiles equivariance. *arXiv*, 2010.

[139] S. B. Hrdy. Evolutionary context of human development: The cooperative breeding model. In C.S. Carter, L. Ahnert, K.E. Grossmann, S.B. Hrdy, M.E. Lamb, S.W. Porges, and N. Sachser, editors, *Attachment and Bonding: A New Synthesis*, Dahlem workshop reports, pages 9–32. MIT Press, 2005.

[140] Kazumoto Iguchi and Hiroaki Yamada. Exactly solvable scale-free network model. *Physical Review E*, 71(3):036144, 2005.

[141] Matthew O Jackson. *Social and economic networks*. Princeton university press, 2010.

[142] Matthew O Jackson, Tomas Rodriguez-Barraquer, and Xu Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–97, 2012.

[143] Matthew O Jackson and Brian W Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, 2007.

[144] A. V. Jaeggi, J. M. Burkart, and C. P. Van Schaik. On the psychology of cooperation in humans and other primates: combining the natural history and experimental evidence of prosociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2723–2735, 2010.

[145] Adrian V Jaeggi and Michael Gurven. Natural cooperators: food sharing in humans and other primates. *Evolutionary Anthropology: Issues, News, and Reviews*, 22(4):186–195, 2013.

[146] Adrian V Jaeggi and Michael Gurven. Reciprocity explains food sharing in humans and other primates independent of kin selection and tolerated scrounging: a phylogenetic meta-analysis. *Proceedings of the Royal Society B: Biological Sciences*, 280(1768):20131615, 2013.

[147] Almerima Jamakovic, Priya Mahadevan, Amin Vahdat, Marián Boguná, and Dmitri Krioukov. How small are building blocks of complex networks. *arXiv preprint arXiv:0908.1143*, 2009.

[148] Hawoong Jeong, Zoltan Néda, and Albert-László Barabási. Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)*, 61(4):567, 2003.

[149] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[150] Samuel Johnson, Joaquín J Torres, J Marro, and Miguel A Munoz. Entropic origin of disassortativity in complex networks. *Physical review letters*, 104(10):108702, 2010.

[151] H. Kaplan and K. Hill. Food sharing among ache foragers: Tests of explanatory hypotheses. *Current anthropology*, 26(2):223–246, 1985.

[152] H. S. Kaplan, P. L. Hooper, and M. Gurven. The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3289–3299, 2009.

[153] Hillard Kaplan, Michael Gurven, Kim Hill, and A Magdalena Hurtado. The natural history of human food sharing and cooperation: a review and a new multi-individual approach to the negotiation of norms. In Herbert Gintis, Samuel Bowles, Robert T Boyd, and Ernst Fehr, editors, *Moral sentiments and material interests: The foundations of cooperation in economic life*, volume 6, pages 75–113. MIT press, 2005.

[154] Hillard Kaplan, Kim Hill, Jane Lancaster, and A Magdalena Hurtado. A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 9(4):156–185, 2000.

[155] Hillard S Kaplan, Eric Schniter, Vernon L Smith, and Bart J Wilson. Risk and the evolution of human exchange. *Proceedings of the Royal Society B: Biological Sciences*, 279(1740):2930–2935, 2012.

[156] Dean Karlan, Markus Mobius, Tanya Rosenblat, and Adam Szeidl. Trust and social collateral. *The Quarterly Journal of Economics*, 124(3):1307–1361, 2009.

[157] Richard M Karp. The transitive closure of a random digraph. *Random Structures & Algorithms*, 1(1):73–93, 1990.

[158] Claudia Kasper and Monique Borgerhoff Mulder. Who helps and why? cooperative networks in mpimbwe. *Current Anthropology*, 56(5):701–732, 2015.

[159] Konstantinos V Katsikopoulos, Lael J Schooler, and Ralph Hertwig. The robust beauty of ordinary information. *Psychological review*, 117(4):1259, 2010.

[160] K. Kayser and D. Armbruster. Social optima of need-based transfers. *Physica A: Statistical Mechanics and its Applications*, 536:121011, 2019.

[161] Qing Ke and Yong-Yeol Ahn. Tie strength distribution in scientific collaboration networks. *Physical Review E*, 90(3):032804, 2014.

[162] Lawrence H Keeley. Hunter-gatherer economic complexity and "population pressure": A cross-cultural analysis. *Journal of anthropological archaeology*, 7(4):373–411, 1988.

[163] Susan Kent. Sharing in an egalitarian kalahari community. *Man*, pages 479–514, 1993.

[164] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[165] Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S Tomkins. The web as a graph: measurements, models, and methods. In *International Computing and Combinatorics Conference*, pages 1–17. Springer, 1999.

[166] Konstantin Klemm and Victor M Eguiluz. Highly clustered scale-free networks. *Physical Review E*, 65(3):036123, 2002.

[167] Dmitry Kobak and Philipp Berens. The art of using t-sne for single-cell transcriptomics. *Nature communications*, 10(1):1–14, 2019.

[168] Tamara G Kolda, Ali Pinar, Todd Plantenga, and Comandur Seshadhri. A scalable generative graph model with community structure. *SIAM Journal on Scientific Computing*, 36(5):C424–C452, 2014.

[169] Jeremy Koster. Interhousehold meat sharing among mayangna and miskito horticulturalists in nicaragua. *Human Nature*, 22(4):394–415, 2011.

[170] Karen L Kramer and Peter T Ellison. Pooled energy budgets: Resituating human energy-allocation trade-offs. *Evolutionary Anthropology: Issues, News, and Reviews*, 19(4):136–147, 2010.

[171] Paul L Krapivsky, Sidney Redner, and Francois Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85(21):4629, 2000.

[172] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[173] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D Sivakumar, Andrew Tomkins, and Eli Upfal. Stochastic models for the web graph. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 57–65. IEEE, 2000.

[174] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the web for emerging cyber-communities. *Computer networks*, 31(11):1481–1493, 1999.

[175] Sanjay Kumar, Dipti Lohia, Darsh Pratap, Ashutosh Krishna, and BS Panda. Mder: modified degree with exclusion ratio algorithm for influence maximisation in social networks. *Computing*, 104(2):359–382, 2022.

[176] Jérôme Kunegis. Konect: the koblenz network collection. In *Proceedings of the 22nd international conference on world wide web*, pages 1343–1350, 2013.

[177] Hsiang-Tsung Kung, Fabrizio Luccio, and Franco P Preparata. On finding the maxima of a set of vectors. *Journal of the ACM (JACM)*, 22(4):469–476, 1975.

[178] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

[179] Michael D Lee and Shunan Zhang. Evaluating the coherence of take-the-best in structured environments. *Judgment and Decision Making*, 7(4):360, 2012.

[180] D. L. Leonetti and B. Chabot-Hanowell. The foundation of kinship. *Human Nature*, 22(1):16–40, 2011.

[181] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11(Feb):985–1042, 2010.

[182] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM, 2005.

[183] Michael Ley. The dblp computer science bibliography: Evolution, research issues, perspectives. In *International symposium on string processing and information retrieval*, pages 1–10. Springer, 2002.

[184] Lun Li, David Alderson, John C Doyle, and Walter Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4):431–523, 2005.

[185] Lun Li, David Alderson, Walter Willinger, and John Doyle. A first-principles approach to understanding the internet's router-level topology. *ACM SIGCOMM computer communication review*, 34(4):3–14, 2004.

[186] Menghui Li, Jinshan Wu, Dahui Wang, Tao Zhou, Zengru Di, and Ying Fan. Evolving model of weighted networks inspired by scientific collaboration networks. *Physica A: Statistical Mechanics and its Applications*, 375(1):355–364, 2007.

[187] Steffen Lippert and Giancarlo Spagnolo. Networks of relations and word-of-mouth communication. *Games and Economic Behavior*, 72(1):202–217, 2011.

[188] Nelly Litvak, Werner RW Scheinhardt, and Yana Volkovich. In-degree and pagerank: why do they follow similar power laws? *Internet mathematics*, 4(2-3):175–198, 2007.

[189] Wei-Yin Loh. Improving the precision of classification trees. *The Annals of Applied Statistics*, pages 1710–1737, 2009.

[190] Sergi Lozano, Alex Arenas, and Angel Sanchez. Mesoscopic structure conditions the emergence of cooperation on social networks. *PLoS one*, 3(4):e1892, 2008.

[191] Pádraig Mac Carron, Kimmo Kaski, and Robin Dunbar. Calling dunbar's numbers. *Social Networks*, 47:151–155, 2016.

[192] Benoit Mandelbrot. An informational theory of the statistical structure of language. *Communication theory*, 84:486–502, 1953.

[193] Massimo Marchiori and Vito Latora. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*, 285(3-4):539–546, 2000.

[194] Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

[195] J. P. McFall. Rational, normative, descriptive, prescriptive, or choice behavior? the search for integrative metatheory of decision making. *Behavioral Development Bulletin*, 20(1):45, 2015.

[196] John McMillan and Christopher Woodruff. Interfirm relationships and informal credit in vietnam. *The Quarterly Journal of Economics*, 114(4):1285–1320, 1999.

[197] J Miller McPherson and Lynn Smith-Lovin. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American sociological review*, pages 370–379, 1987.

[198] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[199] Robert K Merton et al. The matthew effect in science. *Science*, 159(3810):56–63, 1968.

[200] Andrea B Migliano, Abigail E Page, Jesus Gómez-Gardeñes, Gul Deniz Salali, Sylvain Viguier, Mark Dyble, James Thompson, Nikhill Chaudhary, Daniel Smith, Janis Strods, et al. Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour*, 1(2):1–6, 2017.

[201] Andrea Bamberg Migliano and Lucio Vinicius. The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philosophical Transactions of the Royal Society B*, 377(1843):20200317, 2022.

[202] Dina Mistry, Qian Zhang, Nicola Perra, and Andrea Baronchelli. Committed activists and the reshaping of status-quo social consensus. *Physical Review E*, 92(4):042805, 2015.

[203] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[204] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[205] James Moody. Peer influence groups: identifying dense clusters in large networks. *Social networks*, 23(4):261–283, 2001.

[206] M. E. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[207] Mark Newman. *Networks*. Oxford university press, 2018.

[208] Mark EJ Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.

[209] Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409, 2001.

[210] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[211] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

[212] Mark EJ Newman and Juyong Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.

[213] D. A. Nolin. Food-sharing networks in lamalera, indonesia: status, sharing, and signaling. *Evolution and Human Behavior*, 33(4):334–345, 2012.

[214] David A Nolin. Food-sharing networks in lamalera, indonesia. *Human Nature*, 21(3):243–268, 2010.

[215] Samir Okasha. *Evolution and the levels of selection*. Clarendon Press, 2006.

[216] Liudmila Ostroumova, Alexander Ryabchenko, and Egor Samosvat. Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 185–202. Springer, 2013.

[217] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 435(7043):814–818, 2005.

[218] Pradumn Kumar Pandey and Bibhas Adhikari. A parametric model approach for structural reconstruction of scale-free networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2072–2085, 2017.

[219] Juyong Park and Mark EJ Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, 2004.

[220] Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and correlation properties of the internet. *Physical review letters*, 87(25):258701, 2001.

[221] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[222] John Q Patton. Meat sharing for coalitional support. *Evolution and human behavior*, 26(2):137–157, 2005.

[223] Mateusz Pawlik and Nikolaus Augsten. Tree edit distance: Robust and memory-efficient. *Information Systems*, 56:157–173, 2016.

[224] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[225] Tiago P Peixoto. Efficient monte carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, 89(1):012804, 2014.

[226] Tiago P. Peixoto. The graph-tool python library. *figshare*, 2014.

[227] María Pereda, Débora Zurro, José I Santos, Ivan Briz i Godino, Myrian Álvarez, Jorge Caro, and José M Galán. Emergence and evolution of cooperation under resource pressure. *Scientific reports*, 7(1):1–10, 2017.

[228] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2(1):1–7, 2012.

[229] Seth Pettie. On the comparison-addition complexity of all-pairs shortest paths. In *International Symposium on Algorithms and Computation*, pages 32–43. Springer, 2002.

[230] Seth Pettie and Vijaya Ramachandran. Computing shortest paths with comparisons and additions. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 267–276, 2002.

[231] Francisco Plana. Quickcent paper data and code. *figshare*, 2023.

[232] Francisco Plana and Jorge Pérez. Quickcent: A fast and frugal heuristic for centrality estimation on networks. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 238–245. IEEE, 2018.

[233] Francisco Plana and Jorge Pérez. Food-sharing data and code. *figshare*, 2022.

[234] Leon R Planken, Mathijs M de Weerdt, and Roman PJ van der Krogt. Computing all-pairs shortest paths by leveraging low treewidth. *Journal of artificial intelligence research*, 43:353–388, 2012.

[235] Derek de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306, 1976.

[236] John R Quinlan et al. Learning with continuous classes. In *5th Australian joint conference on artificial intelligence*, volume 92, pages 343–348. Singapore, 1992.

[237] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[238] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

[239] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *science*, 297(5586):1551–1555, 2002.

[240] Jakob Raymaekers and Peter Rousseeuw. A generalized spatial sign covariance matrix. *Journal of Multivariate Analysis*, 171:94–111, 2019.

[241] Elspeth Ready and Eleanor A Power. Why wage earners hunt: food sharing, social structure, and influence in an arctic mixed economy. *Current Anthropology*, 59(1):74–97, 2018.

[242] Sidney Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2):131–134, 1998.

[243] Sidney Redner. Citation statistics from more than a century of physical review. *arXiv preprint physics/0407137*, 2004.

[244] Fu-Xin Ren, Hua-Wei Shen, and Xue-Qi Cheng. Modeling the clustering in citation networks. *Physica A: Statistical Mechanics and its Applications*, 391(12):3533–3539, 2012.

[245] Erik J Ringen, Pavel Duda, and Adrian V Jaeggi. The evolution of daily food sharing: A bayesian phylogenetic analysis. *Evolution and Human Behavior*, 40(4):375–384, 2019.

[246] Matei Ripeanu, Ian Foster, and Adriana Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *arXiv preprint cs/0209028*, 2002.

[247] Javier Rivas. Friendship selection. *International Journal of Game Theory*, 38(4):521–538, 2009.

[248] P. J. Rousseeuw. Tutorial to robust statistics. *Journal of chemometrics*, 5(1):1–20, 1991.

[249] Alejandro F Rozenfeld, Reuven Cohen, Daniel Ben-Avraham, and Shlomo Havlin. Scale-free networks on lattices. *Physical Review Letters*, 89(21):218701, 2002.

[250] David E Rumelhart, James L McClelland, PDP Research Group, et al. *Parallel distributed processing*, volume 1. IEEE, 1988.

[251] Fernando P Santos, Jorge M Pacheco, Ana Paiva, and Francisco C Santos. Structural power and the evolution of collective fairness in social networks. *PloS one*, 12(4), 2017.

[252] Francisco C Santos, Marta D Santos, and Jorge M Pacheco. Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216, 2008.

[253] Benjamin Scheibehenne and Arndt Bröder. Predicting wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, 23(3):415–426, 2007.

[254] Hamed Seyed-Allaei, Ginestra Bianconi, and Matteo Marsili. Scale-free networks with an exponent less than two. *Physical Review E*, 73(4):046113, 2006.

[255] Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[256] Jan Skopek, Florian Schulz, and Hans-Peter Blossfeld. Who contacts whom? educational homophily in online mate selection. *European Sociological Review*, 27(2):180–195, 2011.

[257] D. Smith, M. Dyble, K. Major, A. E. Page, N. Chaudhary, G. D. Salali, and R. Mace. A friend in need is a friend indeed: Need-based sharing, rather than cooperative assortment, predicts experimental resource transfers among agta hunter-gatherers. *Evolution and human behavior*, 40(1):82–89, 2019.

[258] Eric Alden Smith and Rebecca L Bliege Bird. Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and human behavior*, 21(4):245–261, 2000.

[259] Brent Snook, Michele Zito, Craig Bennell, and Paul J Taylor. On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21(1):1–26, 2005.

[260] Ray Solomonoff and Anatol Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biology*, 13(2):107–117, 1951.

[261] Donald F Specht. A general regression neural network. *IEEE transactions on neural networks*, 2(6):568–576, 1991.

[262] Michele Starnini, Mattia Frasca, and Andrea Baronchelli. Emergence of metapopulations and echo chambers in mobile agents. *Scientific reports*, 6:31834, 2016.

[263] Hrvoje Štefančić and Vinko Zlatić. "winner takes it all": Strongest node rule for evolution of scale-free networks. *Physical review E*, 72(3):036105, 2005.

[264] Clara Stegehuis, Remco van der Hofstad, AJEM Janssen, and Johan SH van Leeuwaarden. Clustering spectrum of scale-free networks. *Physical Review E*, 96(4):042309, 2017.

[265] J. R. Stevens and F. A. Cushman. Cognitive constraints on reciprocity and tolerated scrounging. *Behavioral and Brain Sciences*, 27(4):569–570, 2004.

[266] Michael PH Stumpf, Carsten Wiuf, and Robert M May. Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proceedings of the National Academy of Sciences*, 102(12):4221–4224, 2005.

[267] Lovro Šubelj and Marko Bajec. Software systems through complex networks science: Review, analysis and applications. In *Proceedings of the First International Workshop on Software Mining*, pages 9–16, 2012.

[268] Jun Sun, Jérôme Kunegis, and Steffen Staab. Predicting user roles in social networks using transfer learning with feature transformation. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 128–135. IEEE, 2016.

[269] Reiko Tanaka. Scale-rich metabolic networks. *Physical review letters*, 94(16):168101, 2005.

[270] Mario V Tomasello, Nicola Perra, Claudio J Tessone, Márton Karsai, and Frank Schweitzer. The role of endogenous and exogenous mechanisms in the formation of r&d networks. *Scientific reports*, 4(1):1–12, 2014.

[271] Jeffrey Travers and Stanley Milgram. The small world problem. *Phychology Today*, 1:61–67, 1967.

[272] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, pages 425–443, 1969.

[273] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pages 141–162. Springer, 1975.

[274] Christopher Udry. Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies*, 61(3):495–526, 1994.

[275] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[276] Alexei Vazquez. Statistics of citation networks. *arXiv preprint cond-mat/0105031*, 2001.

[277] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):056104, 2003.

[278] Alexei Vázquez, Alessandro Flammini, Amos Maritan, and Alessandro Vespignani. Modeling of protein interaction networks. *Complexus*, 1(1):38–44, 2003.

[279] Fernando Vega-Redondo. Building up social capital in a changing world. *Journal of Economic Dynamics and Control*, 30(11):2305–2338, 2006.

[280] Ivan Voitalov, Pim Van Der Hoorn, Maksim Kitsak, Fragkiskos Papadopoulos, and Dmitri Krioukov. Weighted hypersoft configuration model. *Physical Review Research*, 2(4):043157, 2020.

[281] Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad, and Dmitri Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3):033034, 2019.

[282] Bettina von Helversen and Jörg Rieskamp. The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, 137(1):73, 2008.

[283] C. Von Rueden, M. Gurven, and H. Kaplan. Why do men seek status? fitness payoffs to dominance and prestige. *Proceedings of the Royal Society B: Biological Sciences*, 278(1715):2223–2232, 2011.

[284] Christopher R von Rueden, Daniel Redhead, Rick O'Gorman, Hillard Kaplan, and Michael Gurven. The dynamics of men's cooperation and social status in a small-scale society. *Proceedings of the Royal Society B*, 286(1908):20191367, 2019.

[285] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Cambridge university press, 1994.

[286] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[287] Geoffrey B West, James H Brown, and Brian J Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–126, 1997.

[288] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[289] David Sloan Wilson, Elinor Ostrom, and Michael E Cox. Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization*, 90:S21–S32, 2013.

[290] Bruce Winterhalder. Diet choice, risk, and food sharing in a stochastic environment. *Journal of anthropological archaeology*, 5(4):369–392, 1986.

[291] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, 2nd edition, 2005.

[292] Jan K Woike, Ulrich Hoffrage, and Ralph Hertwig. Estimating quantities: Comparing simple heuristics and machine learning algorithms. In *Artificial Neural Networks and Machine Learning–ICANN 2012: 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11-14, 2012, Proceedings, Part II 22*, pages 483–490. Springer, 2012.

[293] Markus Wübben and Florian v Wangenheim. Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, 72(3):82–93, 2008.

[294] Qi Xuan, Yanjun Li, and Tie-Jun Wu. Growth model for complex networks with hierarchical and modular structures. *Physical Review E*, 73(3):036105, 2006.

[295] G Udny Yule. A mathematical theory of evolution, based on the conclusions of dr. jc willis, frs. *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character*, 213:21–87, 1925.

[296] Gorka Zamora-López, Vinko Zlatić, Changsong Zhou, Hrvoje Štefančić, and Jürgen Kurths. Reciprocity of networks with degree correlations and arbitrary degree sequences. *Physical Review E*, 77(1):016106, 2008.

[297] Linjun Zhang, Michael Small, and Kevin Judd. Exactly scale-free scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 433:182–197, 2015.

[298] Bin Zhou, Xiangyi Meng, and H Eugene Stanley. Power-law distribution of degree–degree distance: A better representation of the scale-free property of complex networks. *Proceedings of the national academy of sciences*, 117(26):14812–14818, 2020.

[299] Tao Zhou, Gang Yan, and Bing-Hong Wang. Maximal planar networks with large clustering coefficient and power-law degree distribution. *Physical Review E*, 71(4):046141, 2005.

Annex A

Fragmentation threshold under random removal

In a study [7] described at Section 2.2.1, it is reported that for scale-free networks, no threshold for fragmentation is observed, that is, the size of the largest cluster slowly decreases as the removed node proportion p increases. In contrary, for Erdös-Rényi graphs, under random node removal of a fraction p of nodes, there is a critical proportion $p = p_c$ of nodes at which the main cluster of nodes breaks into small pieces, and the relative size of the largest cluster tends quickly to zero. This phenomenon was analytically studied in a work by Cohen et al (2000) [70]. In this section we review their argument, to highlight the importance of the moments of the power-law distribution in explaining some of its features.

The argument starts by stating that, for a giant component to exist -in an undirected graph-, each node belonging to it must be connected to, at least, two other nodes on average. Thus, denoting by $\mathbb{P}(k_i|i \leftrightarrow j)$ the conditional probability that node i has degree k_i, given that it is connected to node j, we can write the condition for the existence of a giant component as follows

$$\langle k_i|i \leftrightarrow j\rangle = \sum_{k_i} k_i \mathbb{P}(k_i|i \leftrightarrow j) = 2, \tag{A.1}$$

where the angular brackets denote an ensemble average. Now, by Bayes theorem,

$$\mathbb{P}(k_i|i \leftrightarrow j) = \frac{\mathbb{P}(i \leftrightarrow j|k_i)\mathbb{P}(k_i)}{\mathbb{P}(i \leftrightarrow j)}, \tag{A.2}$$

and the terms in this expression, hold that $\mathbb{P}(i \leftrightarrow j|k_i) = \frac{1}{n-1} \cdot k_i$, and $\mathbb{P}(i \leftrightarrow j) = \frac{|E|}{n(n-1)/2} = \frac{\langle k\rangle}{n-1}$, where n is the number of nodes of the graph, $|E|$ is the number of edges, and $\langle k\rangle$ is the average degree. Replacing these last expressions on (A.2) and on (A.1), it follows that the criterion (A.1) is equivalent to

$$\frac{\langle k^2\rangle}{\langle k\rangle} = 2, \tag{A.3}$$

which is a criterion independent of the particular degree distribution. This argument is valid if loops are neglected. The probability of an edge to form a loop is indeed negligible [70].

Consider now a random removal of a fraction p of the nodes. This will alter the connectivity distribution of a node. Indeed, consider a node with initial connectivity k_0, chosen from an initial distribution $P(k_0)$. After a random breakdown of size p, this node has now k edges left with probability $\binom{k_0}{k}(1-p)^k p^{k_0-k}$, since (k_0-k) edges are lost, each with probability p, and k edges remain, each with probability $(1-p)$. Therefore, the probability $P'(k)$ of having a node of degree k in the new network is given by

$$P'(k) = \sum_{k_0 \geq k}^{\infty} P(k_0) \binom{k_0}{k} (1-p)^k p^{k_0-k}.$$

We want now to compute the moments $\langle k \rangle'$, $\langle k^2 \rangle'$ of this new distribution. Thus,

$$\begin{aligned}
\langle k \rangle' &= \sum_{i=0}^{\infty} i P'(i) \\
&= \sum_{i=0}^{\infty} \sum_{k_0 \geq i}^{\infty} i P(k_0) \binom{k_0}{i} (1-p)^i p^{k_0-i} \\
&= (1-p) \sum_{k_0=0}^{\infty} k_0 P(k_0) \underbrace{\sum_{i=0}^{k_0} \binom{k_0-1}{i-1} (1-p)^{i-1} p^{k_0-i}}_{=1} \\
&= (1-p)\langle k_0 \rangle,
\end{aligned}$$

where $\langle k_0 \rangle$ is the mean of the initial distribution P, and we used that $\binom{k_0-1}{i-1} = \binom{k_0}{i} - \binom{k_0-1}{i}$. In order to compute the second moment, we will use the following equality $\langle k^2 \rangle' = \langle k^2 - k + k \rangle' = \langle k(k-1) \rangle' + \langle k \rangle'$. Thus, we must compute

$$\begin{aligned}
\langle k(k-1) \rangle' &= \sum_{i=0}^{\infty} i(i-1) P'(i) \\
&= \sum_{i=0}^{\infty} \sum_{k_0 \geq i}^{\infty} i(i-1) P(k_0) \binom{k_0}{i} (1-p)^i p^{k_0-i} \\
&= (1-p)^2 \sum_{k_0=0}^{\infty} k_0(k_0-1) P(k_0) \cdot \\
&\qquad \underbrace{\sum_{i=0}^{k_0} \binom{k_0-2}{i-2} (1-p)^{i-2} p^{k_0-i}}_{=1} \\
&= (1-p)^2 \langle k_0(k_0-1) \rangle,
\end{aligned}$$

which, in turn, allows us to compute

$$\begin{aligned}\langle k^2\rangle' &= \langle k(k-1)\rangle' + \langle k\rangle' \\ &= (1-p)^2\langle k_0(k_0-1)\rangle + (1-p)\langle k_0\rangle \\ &= (1-p)^2\langle k_0^2\rangle - (1-p)^2\langle k_0\rangle + (1-p)\langle k_0\rangle \\ &= (1-p)^2\langle k_0^2\rangle + p(1-p)\langle k_0\rangle.\end{aligned}$$

With this in hand, the criterion (A.3) can be applied to the moments $\langle k\rangle'$, $\langle k^2\rangle'$ -or in other words, to the new distribution $P'(k)$ resulting from the removal of a fraction p of nodes-. That is, the criterion (A.3) may be expressed as

$$\frac{\langle k_0^2\rangle}{\langle k_0\rangle}(1-p_c) + p_c = 2, \tag{A.4}$$

and solving the expression for p_c, (A.4) gives

$$p_c = 1 - \frac{1}{\frac{\langle k_0^2\rangle}{\langle k_0\rangle} - 1}.$$

This is an expression for the critical proportion of removed nodes up to which there is a giant component, in function of the moments of the original distribution. This expression is independent of the particular degree distribution, and so applicable to all randomly connected networks.

In particular, for an Erdös-Rényi graph, with a Poisson degree distribution of mean $\langle k_0\rangle$, it holds that $\langle k_0^2\rangle = \langle k_0\rangle(1 + \langle k_0\rangle)$ -since the variance of a Poisson distribution is equal to its mean-, and hence, the critical proportion is given by $p_c = 1 - \frac{1}{\langle k_0\rangle}$. This is a finite value becoming greater if the network is denser. On the other hand, a Power-law degree distribution of exponent $\alpha \in (2, 3]$, as most networks from Table 2.1, has a finite mean but a divergent second moment, as seen at Section 2.2.5. In this case, $p_c \approx 1$, which is consistent to the results reported at Albert, Jeong and Barabási (2000) [7].

Annex B

Supplementary Information for "QuickCent: a fast and frugal heuristic for harmonic centrality estimation on scale-free networks"

Abstract

This annex corresponds to the online supplementary material for the paper *QuickCent: a fast and frugal heuristic for harmonic centrality estimation on scale-free networks* submitted to *Computing* journal. It includes the practical procedure used to estimate the lower limit of the power-law distribution, and the experiments for checking the fulfillment of assumptions of QuickCent by the distinct networks shown in the Results section of Chapter 3.

B.1 Practical procedure to estimate the lower limit of the power-law distribution

Required background of the power-law distribution is introduced on Sections 2.2.3, 2.2.5 and 2.3.3 from the survey chapter. Methods to estimate the parameters of the power-law are shown in Section 2.3.3. In this section we explain the practical procedure used to estimate the lower limit $x_{\min}$ of the distribution.

We estimated $x_{\min}$ with the bootstrap method implemented by the *poweRlaw* R package [111], where several samples $x_{\min}$ are drawn and that minimizing $D(x_{\min})$ is selected. We noticed in our experiments that, with high frequency, this method selects $x_{\min}$ as a point with a high value, that is, a $x_{\min}$ value that discards a high portion of the distribution. In the case of empirical networks shown in Section 3.4.4, we have taken the heuristic approach of limiting the search space by an upper bound given by the percentile 20 of the distribution of positive centrality values[1], since we have seen for many datasets this is enough to span the point where the log-log plot of the complementary ECDF starts to behave like a straight line. Other authors giving implementations of this method have also noticed the difficulties when estimating $x_{\min}$[2]. This method has a statistical consistency that has been proved only for some heavy-tailed models [87]. There are alternative methods to optimize the KS statistic that perform, for example, a grid search over a predefined set of exponent values for each possible $x_{\min}$ that, however, have been claimed to present many drawbacks [281].

[1] This is the domain where the power-law fit can be computed.

[2] The commented code from https://github.com/keflavich/plfit says: *..."The MLE for the power-law alpha is very easy to derive given knowledge of the lowest value at which a power law holds, but that point is difficult to derive and must be acquired iteratively."*

B.2 Synthetic networks setting and assumptions verification

The synthetic network model we chose for our simulations is the preferential attachment (PA) growth reviewed in Section 2.3.4, including a small additive constant to the probability of acquiring links that allow isolated nodes to attract arcs[3] [171, 85, 20]. This is a very simple model that reproduces the power-law degree distribution observed in real networks coming from a wide range of fields [20], and provides a convenient setting for studying our heuristics on several prototypical distributions. With the goal of illustrating these distinct patterns, we have plotted in Figure B.1 the harmonic centrality and in-degree distributions for randomly generated graphs of the model from Section 2.3.4 for exponents $0.5, 1, 1.5$.

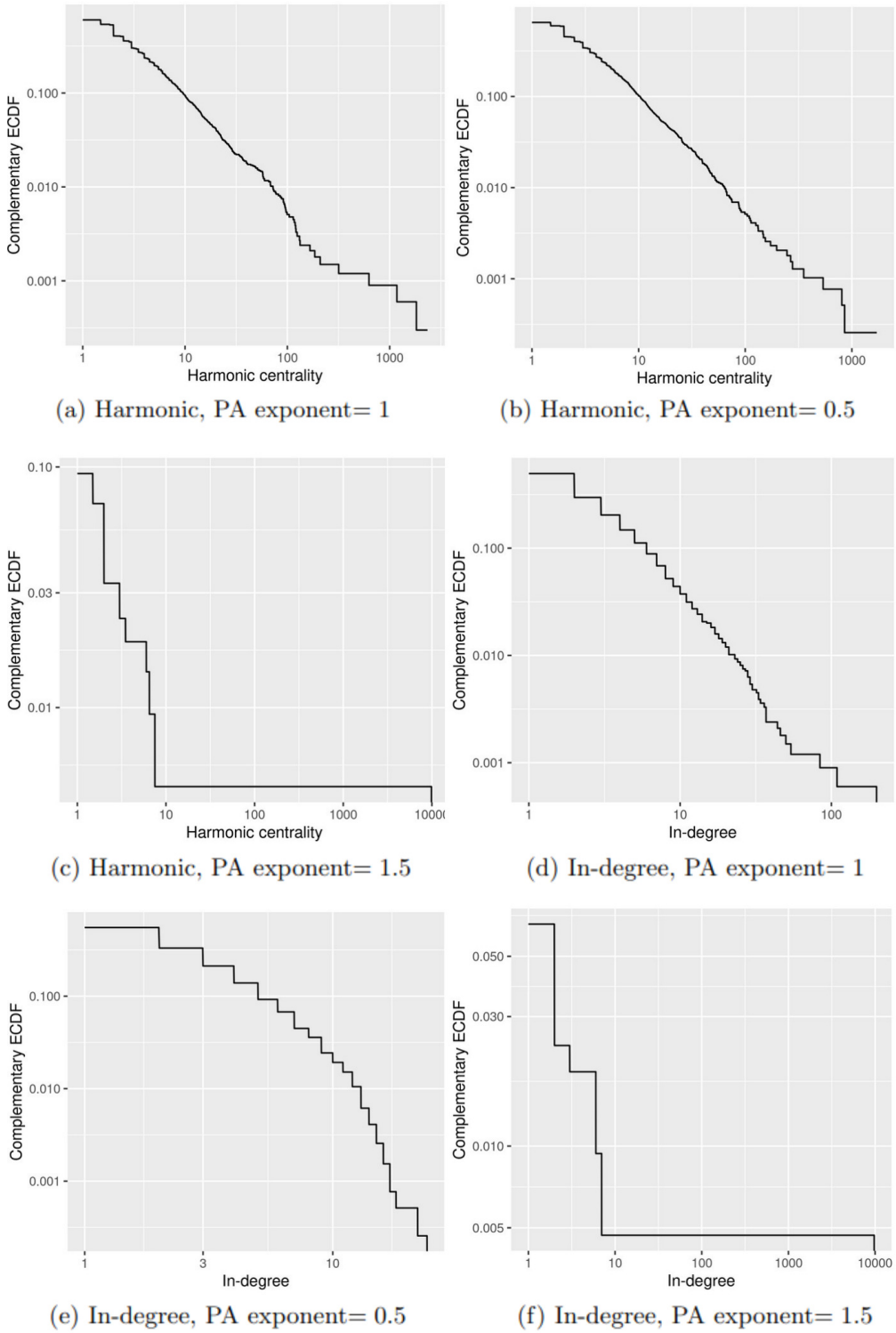
In Figure B.1 (d), (e) and (f), we can visualize the respective patterns for the in-degree generated by PA described in Section 2.3.4. That is, a power-law distribution, a stretched exponential reflected in the scale on the horizontal axis, and the *gelation* regime of a single gel connected to every node reflected in a heavy-tailed distribution even on a log-log plot. On the other hand, in Figure B.1 (a), (b) and (c) the respective patterns of the harmonic centrality are depicted, where the gelation pattern is also obtained for exponent 1.5, and the scale-free behavior is observed for both exponents $1, 0.5$, which is noteworthy. There are works in the literature showing that the in-degree and PageRank centrality of a digraph obey a power-law with the same exponent [188], but we have no knowledge of results describing the distribution of harmonic centrality on digraphs.

Before we can proceed with applying QuickCent, we need to ensure that its assumptions, namely the power-law distribution of the centrality and the monotonic relation between the degree and the centrality, are satisfied by the PA model. We checked the assumption of the power-law distribution of the centrality by using the goodness-of-fit test proposed by Clauset et al. [67]. This test is based on the KS distance d, see Equation (2.12), between the distribution of the empirical data and the fitted power-law model, and produces a p-value p, computed via a Monte Carlo procedure, which estimates the probability that the distance for any random sample is larger than d. Therefore, if p is close to 1, the fit is acceptable since the difference between the empirical data and the model fit can be explained by random fluctuations; otherwise, if p is close to 0, the model is not an appropriate fit to the data [67].

The experiment executed to check the power-law assumption of harmonic centrality was to instantiate a PA network and compute the following magnitudes: an estimate $\hat{x}_{\min}$ of the distribution lower limit via minimization of Equation (2.12), the respective exponent $\hat{\alpha}$ by using Equation (2.11), and the significance of this fit[4]. We also compute the value $\hat{\alpha}_1$ which is an estimation of the exponent α associated to value $x_{\min} = 1$. The objective of computing $\hat{\alpha}_1$ was to verify the error obtained by working with a fixed value of $x_{\min}$ (equal to 1 in this case), which may also tell about the goodness of the power-law fit. The described calculations were repeated over $1,000$ networks. Some descriptive statistics for the different magnitudes computed are shown in Table B.1 (Q25 and Q75 are the 25-th and 75-th

[3]This setting corresponds to the default parameters of the method *sample_pa()* of package *igraph* [76] for the R language.

[4]This test requires the values tested to be strictly positive, so the zero harmonic values present in the sample were discarded for this test.



(a) Harmonic, PA exponent= 1

(b) Harmonic, PA exponent= 0.5

(c) Harmonic, PA exponent= 1.5

(d) In-degree, PA exponent= 1

(e) In-degree, PA exponent= 0.5

(f) In-degree, PA exponent= 1.5

Figure B.1: **Complementary ECDF of harmonic centrality and in-degree for prototypical preferential attachment networks.** Each plot shows with logarithmic (base 10) axes the complementary empirical cumulative density function of the harmonic centrality, and in-degree of randomly generated networks of each representative exponent of the PA model. The distributions of in-degree are known from the literature [171], but the distributions for harmonic centrality are a new result. For PA exponents 1 and 0.5, harmonic is a power-law, and for the gelation in-degree (there is a super-hub) regime, harmonic has an analogous behavior.

	$\hat{x}_{\min}$	$\hat{\alpha}$	**p-value**	$\hat{\alpha} - \hat{\alpha}_1$	**Spearman**	**Spearman p-value**
Q25	6.333	2.135	0.333	-0.027	0.923	0
Median	7.333	2.167	0.582	0.003	0.925	0
Mean	8.343	2.171	0.555	0.008	0.925	0
Q75	9.166	2.203	0.808	0.041	0.928	0

Table B.1: **Results of experiments testing QuickCent's assumptions for PA with exponent 1.** Fields correspond to: fitted lower limit and exponent of the power-law, KS-based p-value of this fit, difference between the fitted exponent and the exponent assuming $x_{\min} = 1$, the Spearman correlation between the logarithms of centrality and in-degree, and its significance. The number of decimal places is truncated to three with respect to the source.

	$\hat{x}_{\min}$	$\hat{\alpha}$	**p-value**	$\hat{\alpha} - \hat{\alpha}_1$	**Spearman**	**Sp. p-value**
Q25	2.500	1.429	0.056	-5.438	0.853	1.043e-82
Median	3.500	1.521	0.274	-4.518	0.882	1.645e-70
Mean	93.351	1.670	0.346	-4.610	0.876	5.647e-29
Q75	4.208	1.774	0.620	-3.650	0.903	3.977e-60

Table B.2: **Results of experiments testing QuickCent's assumptions for PA with exponent 1.5.** Fields correspond to: fitted lower limit and exponent of the power-law, KS-based p-value of this fit, difference between the fitted exponent and the exponent assuming $x_{\min} = 1$, the Spearman correlation between the logarithms of centrality and in-degree, and its significance. The number of decimal places is truncated to three with respect to the source.

percentiles). Estimations were computed with the *poweRlaw* R package [111].

From this table, we can see that 75% of the $1,000$ repetitions have a p-value greater than 0.333, which is greater than the rule-of-thumb threshold of 0.1 to rule out the power law [67]. Hence, the fit given by $\hat{x}_{\min}$ and $\hat{\alpha}$ is considered acceptable and the first assumption is satisfied. In the table, it can be appreciated that the estimated exponents $\hat{\alpha}$ are relatively stable for different values of $\hat{x}_{\min}$, and similar to $\hat{\alpha}_1$, in the sense that their difference is less than the critical amount of one integer which changes the behavior of the moments (see Section 2.2.5 from the survey chapter). Given this observation, in the experiments of Sections 3.4.1, 3.4.2 and 3.4.3 (in the case of the loss of the monotonic map) of the main chapter where QuickCent is applied to estimate centrality on synthetic PA networks, we consider a fixed value of $x_{\min} = 1$. This further simplifies QuickCent training without incurring a big error in the parameter estimates.

To check the assumption of monotonic relationship, in the same simulations performed before, we computed the Spearman correlation between the logarithm of the harmonic centrality and the logarithm of the in-degree[5]. Some statistics for the distribution of correlation values, and the distribution of significant p-values can be reviewed in Table B.1. Here, we see that most of the correlations are almost perfect and significant. Since there is a linear

[5]Here we also removed the zero harmonic values since $\log(0) = -\infty$.

	$\hat{x}_{\min}$	$\hat{\alpha}$	**p-value**	$\hat{\alpha} - \hat{\alpha}_1$	**Spearman**	**Sp. p-value**
Q25	6.358	2.228	0.317	0.188	0.923	0
Median	7.333	2.252	0.540	0.214	0.926	0
Mean	8.373	2.255	0.544	0.216	0.926	0
Q75	9.166	2.280	0.792	0.244	0.928	0

Table B.3: **Results of experiments testing QuickCent's assumptions for PA with exponent 0.5.** Fields correspond to: fitted lower limit and exponent of the power-law, KS-based p-value of this fit, difference between the fitted exponent and the exponent assuming $x_{\min} = 1$, the Spearman correlation between the logarithms of centrality and in-degree, and its significance. The number of decimal places is truncated to three with respect to the source.

relation between the logarithms of the two variables, we can conclude that there is a good exponential fit of the in-degree to the centrality, so the second assumption is satisfied.

As an example for comparison, in Table B.2 the same statistics are reported for the PA model with exponent 1.5. In this case, we see that while the p-values are not low in the central tendency, they are lower than those with exponent 1 and there is an important proportion of them with values lower than the rule-of-thumb of 0.1. There is also an important difference between the general fitted α and that obtained with $x_{\min} = 1$, which may be associated with the different slopes observed in the ECDF plot for lower and higher centrality values. This may also explain the greater variability observed in the values of $x_{\min}$, where the mean shows the presence of higher values. A similar pattern can be seen regarding the correlation between centrality and degree, in the sense of being lower than that of the previous case. The statistics for the exponent 0.5 of PA growth can be reviewed in Table B.3, and are very similar to those obtained with exponent 1. However, in this case, there are higher differences between $\hat{\alpha}$, associated with $\hat{x}_{\min}$, and $\hat{\alpha}_1$, which in fact will have an effect on the robustness experiments reviewed in Section B.3.

B.3 Robustness of estimates

We present here the results of the following robustness experiment. We consider an instance of a PA graph of $10,000$ nodes and created distinct QuickCent models with varying information amounts from the network. This information is extracted from uniform node samples and the formula from (2.11) to estimate the α exponent, with sample sizes (value m in the formula) ranging from 10% to 100% of the total vertex set. Note that in this formula, there will be a number of nodes discarded from the sample depending on the value of $x_{\min}$. From the experiments shown in Section B.2, we chose to use $x_{\min} = 1$ for the experiments in Sections 3.4.1, 3.4.2 and 3.4.3 (in the case of the loss of the monotonic map) of the main chapter using synthetic PA networks, since there is not a big penalty for doing so, at least with the exponents 0.5 and 1 with a centrality distribution nearer to a power-law. Then, for each model, we compute a centrality estimate on every vertex and the difference between this estimate and the real value. Finally, we computed from these differences the MAE on the

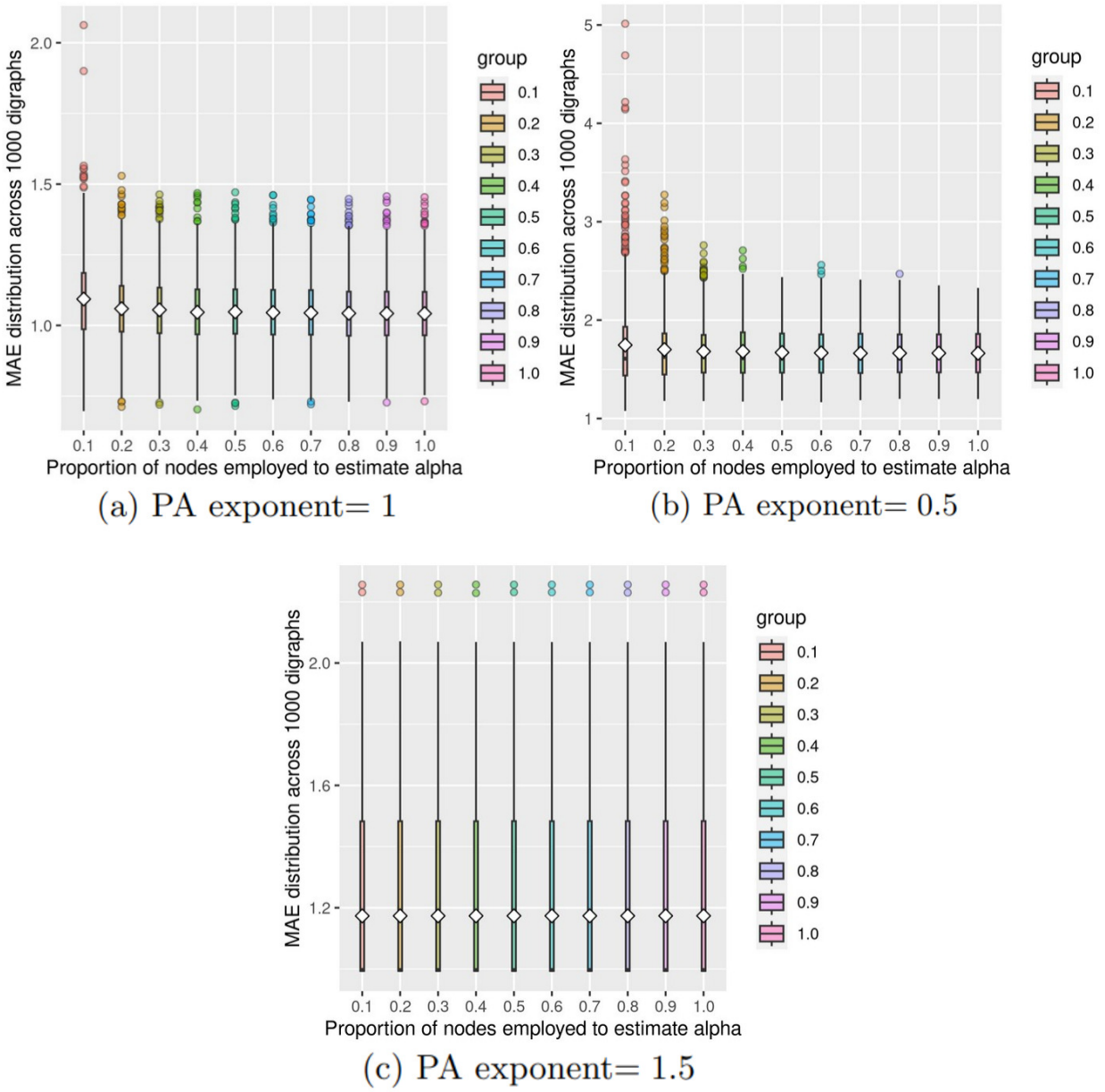


Figure B.2: **Robustness experiments for different exponents of PA digraph instances.** For each sample size there is a boxplot representing the MAE distribution. Each boxplot goes from the 25−th percentile to the 75−th percentile, with a length known as the *inter-quartile range* (IQR). The line inside the box indicates the median, and the rhombus indicates the mean. The whiskers start from the edge of the box and extend to the furthest point within 1.5 times the IQR. Any data point beyond the whisker ends is considered an outlier, and it is drawn as a dot. QuickCent is likely accurate, and with a bound small variance, for all the tested training sizes and PA exponents.

entire digraph and examined the MAE distribution across $1,000$ different instances of PA digraphs. The idea is to assess how the availability of information from the network enhances or deteriorates the quality of the estimates.

Results are shown in Figure B.2. The plots show a performance that is quite accurate and robust to the training size in the executed experiments. It is accurate because the central tendency and most of the distribution contained in the IQR of the MAE errors (see the caption in Figure B.2), does not exceed the two units of difference, taking the average over all network nodes, from the real centrality value, at least for the exponents 1 and 1.5 of PA. Bigger MAE errors appear for the exponent 0.5, however, the most extreme outliers even in this case do not surpass the five units of error. This case reveals that the in-degree does not need to have a scale-free distribution to be functional for QuickCent estimates. Actually, an implicit requirement for the correct working of QuickCent is that this distribution has an injective map from the proportion vector used to the degree quantile values, allowing to distinguish the distinct centrality intervals. For this reason, the length of the proportion vector in these experiments was fixed at 8, since while the accuracy increases with this parameter, for lengthier proportion vectors repeated degree quantile values appear. Thus, the bigger MAE errors for the PA exponent 0.5 come from the exponents $\hat{\alpha}_1$ used in the simulations, which are lower than those $\hat{\alpha}$ associated with the fitted $\hat{x}_{\min}$, see Table B.3. Since they are lower, they predict centrality values that are higher than the real ones, particularly among the highest values.

On the other hand, the performance of QuickCent is robust to the training size in the executed experiments. It is not highly deteriorated for the smaller training sizes of 10 %. The case of exponent 1.5 is special since the performance seems to be quite independent of the training size. This has to do with the fact that there is a structural bias given by the maladjustment between the empirical centrality distribution and the assumed power-law model. However, given that in these simulations $x_{\min} = 1$, the fitted power-law covers the bulk of the distribution composed of low centrality nodes, making the MAE error to be pretty accurate since it is an average over the network. This shows how this measure of the error is favored by low errors committed in a relevant mass of the distribution, as in the power-law case for PA exponents 0.5 and 1.

The feature of making estimates with relatively stable errors could be an advantage in relation to other regression methods, which can suffer potentially a greater impact from scarce data. This issue is addressed in Section 3.4.1 of the main chapter.

B.4 Assumption verification experiments on randomized networks

Table B.4 shows some summary statistics of several variables testing the assumptions to apply QuickCent, on a set of 1000 randomly generated PA networks (exponent 1), subject to degree-preserving randomization of arcs. The p-values show that, even by fixing $x_{\min} = 1$, which is the approximation used by QuickCent, the harmonic centrality distribution is reasonably approximated by a power-law. The fitted parameters of the power-law distribution

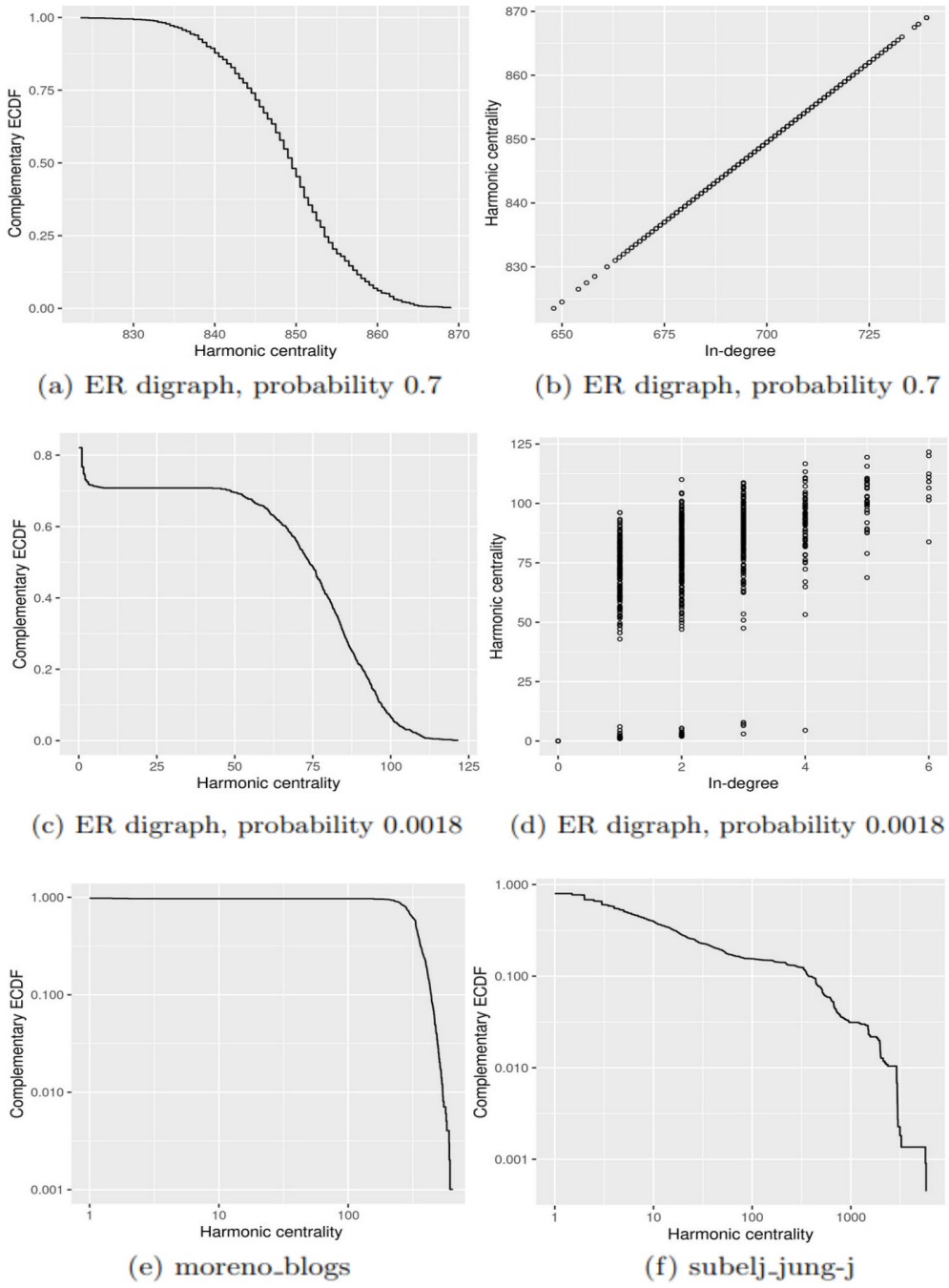
	$\hat{x}_{\min}$	$\hat{\alpha}$	$\hat{\alpha}_1$	**p-value**	**corr_1**	**corr_2**
Q25	2.775	1.858	1.859	0.020	0.918	0.777
Median	3.833	1.960	1.894	0.120	0.927	0.797
Mean	7.636	2.183	1.896	0.269	0.926	0.796
Q75	7.500	2.203	1.932	0.470	0.935	0.817

Table B.4: **Results of experiments testing the assumptions of QuickCent on 1000 PA networks (exponent 1) and degree-preserving randomization.** Fields correspond to: fitted lower limit ($\hat{x}_{\min}$) and power-law exponent($\hat{\alpha}$) of the harmonic centrality, power-law exponent ($\hat{\alpha}_1$) obtained by fixing $x_{\min} = 1$ (the approximation used by QuickCent), the KS-based p-value of this last fit, the Spearman correlation between the positive values of harmonic centrality and in-degree, before (corr_1) and after randomization (corr_2). The number of decimal places is truncated to three with respect to the source. All fields except corr_1 are computed on the network obtained after randomization by swapping the start and end of 10000 randomly selected arc pairs that do not change the in-degree network sequence.

of randomized networks are in fact similar to those shown in Table B.1 corresponding to PA networks without randomization. The two columns with correlation values clearly show the impact of randomization over the map between in-degree and harmonic centrality.

B.5 Sensitivity to connection probability and assumptions verification on Erdös-Rényi digraphs and control networks

Erdös-Rényi (ER) digraphs seem to have a unimodal distribution of harmonic centrality, and an almost perfect correlation between this coefficient and the in-degree, at least for those values of the connection probability p where the mean in-degree $\sim p \cdot N > 1$, corresponding to the values where it is very likely that there is a unique large strongly connected component [157], see Figure B.3. For values of p near the transition point, sets of reachable (or co-reachable) nodes from distinct nodes tend to be smaller, which produces greater variability in the relation mapping in-degree values to harmonic centrality, which lowers the strength of the correlation, see Figure B.3. Table B.5 shows the fitted lower limits and p-values of the power-law fit either on the ER or control digraphs. Each one of the 1000 iterations of fitting over each network is associated with a distinct random seed, impacting for example the bootstrap computations of p-values. In spite of this, for the control networks, the parameters are constant since we are working with the same network, in contrast to the ER digraphs which are newly instanced for each iteration.



(a) ER digraph, probability 0.7 (b) ER digraph, probability 0.7

(c) ER digraph, probability 0.0018 (d) ER digraph, probability 0.0018

(e) moreno_blogs (f) subelj_jung-j

Figure B.3: **Distribution of harmonic centrality (left) and map from in-degree to harmonic (right) on ER, and harmonic centrality in control digraphs.** The plots are generated from random instances of ER digraphs of size $N = 1000$. The top plots represent networks where the probability of connection $p = 0.7$ induces a mean in-degree far greater than 1, the critical value for the existence of a unique giant strong component. This is reflected in a unimodal distribution of harmonic centrality, and a perfect correlation between in-degree and harmonic centrality. The middle plots represent a connection probability $p = 0.0018$ near the transition point, where the correlation is now 0.799. The bottom plots correspond to the harmonic centrality distribution of the two empirical networks used as controls. Plots (a) to (d) show the distinct behaviors of ER digraphs depending on the mean in-degree. Plots (e) and (f) show the harmonic centrality distribution of two empirical networks, more or less near to a power-law.

	$\hat{x}_{\min}^{\text{mb}}$	$\mathbf{p}^{\text{mb}}$	$\hat{x}_{\min}^{\text{sj}}$	$\mathbf{p}^{\text{sj}}$	$\hat{x}_{\min}^{\text{ERmb}}$	$\mathbf{p}^{\text{ERmb}}$	$\hat{x}_{\min}^{\text{ERsj}}$	$\mathbf{p}^{\text{ERsj}}$
Q25	277.416	1	1	1	381.333	0.010	1060.000	0.020
Median	277.416	1	1	1	381.916	0.180	1060.667	0.380
Mean	277.416	1	1	1	381.915	0.347	1060.621	0.431
Q75	277.416	1	1	1	382.500	0.650	1061.167	0.820

Table B.5: **Results of experiments testing the assumptions of QuickCent on ER and control digraphs.** Fields correspond to: fitted lower limit ($\hat{x}_{\min}$) and p-value ($\mathbf{p}$) of this power-law fit to the harmonic centrality distribution of each network. The superscript of each parameter denotes the respective network, 'mb' for moreno_blogs, 'sj' for subelj_jung-j, 'ERmb' for the ER digraph created with the parameters of moreno_blogs, and analogously for 'ERsj'. The number of decimal places is truncated to three with respect to the source.

B.6 Assumptions verification on empirical networks

In Figure B.4 we can see the log-log plot of the complementary empirical cumulative density function of the harmonic centrality in each of the datasets. We can see that all the plots share the feature of having first a flat region of low centrality nodes with high probability, and then there is a turning point starting a new region more similar to a power-law, that however is restricted to a characteristic scale, excluding a possible scale-free behavior. This observation does not exclude this behavior for general datasets, since the observed pattern may well be a consequence of the size of the chosen datasets. In all the plotted datasets, the turning point that would correspond to $\hat{x}_{\min}$ in terms of the power-law distribution, can be visually placed around the percentile 10 to 20 of each distribution, which justifies the use of the upper bound given by the percentile 20 for the search of $\hat{x}_{\min}$ explained in Section B.1. The exception is the plot of the DBLP dataset which, in spite of being very similar to the plots displayed, therein $\hat{x}_{\min}$ may be visually placed around the percentile 60. For this reason, we excluded this dataset in the later analyses. The interested reader can anyway run the provided code [231] to review this dataset analogously to the others.

We performed a similar experiment to that from Section B.2 to study whether the datasets fulfill the assumptions of QuickCent. A p-value to quantify the goodness of fit was computed with the bootstrap procedure from Section B.2. We also computed the Spearman correlation between the logarithms of the in-degree and the centrality[6]. The results are displayed in Table B.6. There we can see that, in general, the datasets present a reasonable fulfillment of the assumptions, the only exception being perhaps the wiki_talk_gl dataset. These results should be taken with care since, from our experimentation, the method for p-value estimation is very sensitive to the value used to bound the search space. The overall exponents obtained agree with the discussion regarding the complementary ECDF plots given in the previous paragraph.

[6]As before, computed on non-zero in-degree and centrality.

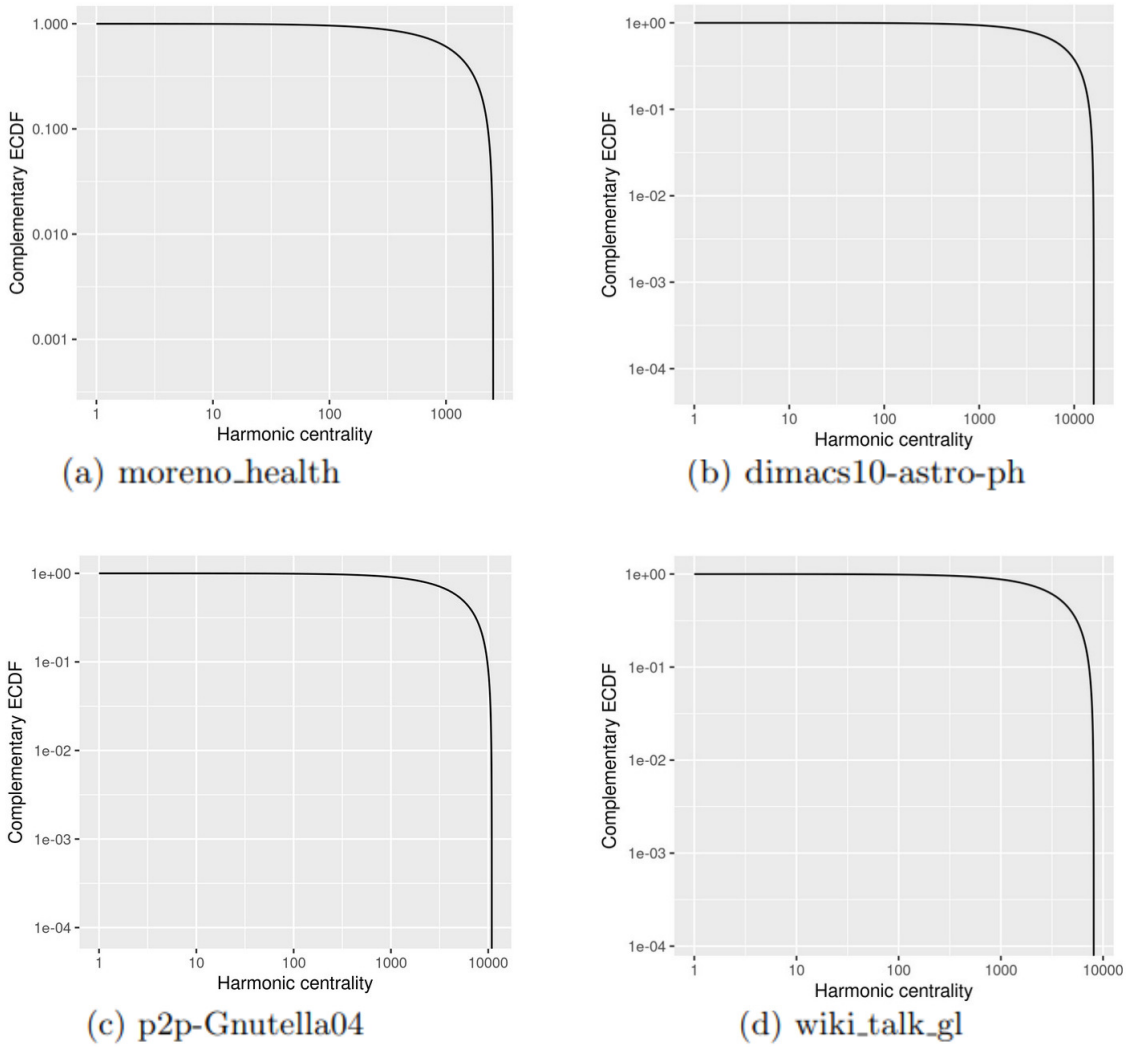


Figure B.4: **Cumulative distribution function of the harmonic centrality on some empirical networks.** Each plot shows, with logarithmic (base 10) axes, the complementary empirical cumulative distribution function of the harmonic centrality in each of the empirical datasets analyzed later. Each plot shows a similar behavior for the studied empirical networks, where the bulk of the distribution are small values, and after some point, starts a power-law behavior with relatively high exponent.

Name	Corr.	p-val	$\hat{x}_{\min}$	$\hat{x}_{\min}$ p-value	$\hat{\alpha}(\hat{x}_{\min})$
moreno_health	0.8	0	339.2662	1	5.519177
dimacs10-astro-ph	0.75	0	2719.642	1	5.199691
p2p-Gnutella04	0.73	0	582.6779	1	4.991946
wiki_talk_gl	0.22	3.5e-83	361.65	0	9.505685

Table B.6: **Indicators of the fulfillment of the assumptions by empirical data sets.** Fields in the table are the network name, its Spearman correlation between the logarithm of the in-degree and the logarithm of the harmonic centrality, the correlation p-value, the fitted lower limit, its corresponding p-value, and the power-law exponent.

Annex C

Food sharing gave birth to social networks

Social networks present distinctive features when compared with other types of networks, particularly the presence of communities, which are subsets of nodes much more densely connected among themselves, than with the rest of the network. In this work, we propose an explanation for this pattern based on the following: groups may be the community solution of hunter-gatherer societies to the survival problem posed by the uncertainty of food. We propose a multi-agent model inspired by a food-sharing dynamic, which combines and formalizes two main notions discussed by some anthropological literature: the reciprocity in the exchanges of food, plus the care for the general welfare of agents. Our preliminary results show that near-to-optimal food-sharing networks exhibit highly-connected groups around special agents that we call hunters, those who inject food into the system. We show the robustness of these results by computer simulations and also by analytical arguments for these simulations.

Food sharing gave birth to **social networks**.

Cohesive groups typically found in social networks may have arisen as a solution by **hunter-gatherer societies** to the survival problem posed by the **uncertainty of food**.

Motivation

A **distinctive attribute** of **social networks** is that they have **communities**, or cohesive groups of agents more densely connected among them than with the rest of network.

A strand in anthropological literature argues about the **central role of food transfers** for the evolution of human foraging niche.

We aim to develop a simple **multi-agent model** where communities appear in food sharing networks optimized to provide **egalitarian access** to food resources under **reciprocal exchanges**

Description of the food-sharing "protocol"

Each **agent** is a **node** in a loop-free **undirected network**. There is a special set of agents that we call **hunters**. Each hunter is able to hunt a prey with fixed probability ph.

We assume that the **prey** is enough to **feed F agents** (including the hunter), where F is a fixed positive integer. The **food** resulting from a hunt is **shared from the hunters** to the rest of the network, only through network edges.

When an **agent receives** *f* units of food, it **consumes a single unit** of food. After this, the agent *a* **chooses** uniformly **at random** one neighbor, say *a0*, and **sends all the remaining food** (*f* −1 units) to *a0*.

Two possible network indices to optimize

If we call the amount of food sent from agent i to agent j as a **flow**

$$f_{i,j} \in \{0, 1, 2, \ldots, F-1\}$$

We define the **reciprocity index** rep(G) of a given network G as the average, over all edges (i,j) of G, of the following,

$$\mathrm{rep}_{i,j}(G) = |\mathbb{E}(f_{i,j}) - \mathbb{E}(f_{j,i})|$$

And also define the **welfare index** W(G) of network G as the standard deviation of the probability of receiving food, over all network nodes.

Model solution

We assume that networks have N nodes. Then, given nonnegative integer values for F, and N, ph $\in [0,1]$, $\alpha \in [0,1]$ and a number NH ≤ N of hunters, we define the following **cost or energy function**,

$$\mathcal{C}(\alpha, G) = \alpha \cdot \mathrm{rep}(G) + (1-\alpha) \cdot \mathrm{wef}(G)$$

The problem to solve in our model is to **find a network** $\tilde{G}_\alpha$ with NH hunters **of minimal cost** $\mathcal{C}(\alpha, \tilde{G}_\alpha)$. The intuition is that we can find networks satisfying either, or both criteria.

Finally, we have used Simulated Annealing algorithm to find local optima networks for this optimization model.

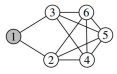
Results for one hunter

Different types of networks are optimal for each or both criteria.

Example for N = 5, ph = 0.2 and F = 5.

Table 1: Costs of candidate networks

Criterion	Complete	Semiclique	Isolated
Reciprocity	0.098	0.26	0
Welfare	0.0253	0.0248	0.08

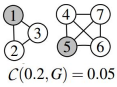


Semiclique network, hunter in gray

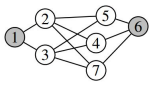
Results for two hunters

Local optima include networks with **well cohesive groups around hunters**, and networks with layers of nodes of equal degree.

Example for N = 7, ph = 0.2 and F = 3.



$\mathcal{C}(0.2, G) = 0.05$



$\mathcal{C}(0.2, G) = 0.036$

dcc CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE CHILE

Instituto Milenio Fundamentos de los datos

Jorge Pérez
Francisco Plana
UChile & IMFD
Contact:
<**franciscoplana**@gmail.com>

Annex D

Supplementary Information for "Modularity of food-sharing networks minimises the risk for individual and group starvation in hunter-gatherer societies"

Abstract

This annex corresponds to the online appendices section for the paper *Modularity of food-sharing networks minimises the risk for individual and group starvation in hunter-gatherer societies*, accepted for publication by *PLOS ONE* journal. It includes the implementation details of several components of the model and employed methods shown in Chapter 4, namely, the estimate of the probability of eating, the structure of the simulations, the evolutionary algorithms, clustering procedures and decision trees.

D.1 Estimate of probability of eating

In this section, we obtain an estimate of the probability of eating for every node, according to the model assumptions and food sharing protocol. With this goal in mind, we introduce some basic graph notions. We can establish an equivalence between a digraph $D = (V, E)$ of N vertices and its *adjacency matrix*,

$$\mathcal{M}(D)_{i,j} = \begin{cases} 1, & (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

and we can define a finite *walk* w in a digraph as a sequence of edges $\{e_i\}_{i=1}^{n-1}$ for which there is a sequence of vertices $\{v_i\}_{i=1}^{n}$, possibly repeated, such that $e_i = (v_i, v_{i+1}), \forall i = 1, \ldots, n-1$. If w is a walk with vertex sequence $\{v_i\}_{i=1}^{n}$, we say that w is a walk *from* v_1 *to* v_n *of length* $n-1$, or a $(n-1)-$walk. It is not difficult to prove that, coefficient $\mathcal{M}(D)_{i,j}^{n}$ from the $n-$th power of $\mathcal{M}(D)$, is equal to the number of walks of length n from i to j in D. We will use this property in what follows. For completeness, we can say that $\mathcal{M}(D)^0 = I_{N\times N}$, the identity matrix, for every network D.

Now, in order to obtain an expression for $pe(v)$, the probability of eating, at each time step, by agent v, let us assume by now that there is just one hunter h to derive an expression for the probability $pe(v, h)$ of v getting feed from the only hunter h. It turns out that it is simpler to write this probability as a function of the negation of the complement event, which corresponds to v not being reached by a $(F-1)-$walk W of nodes sharing a prey from h, or also to the intersection of events $E_i, i = 0, \ldots, F-1$, where each $E_i = \{v_i \neq v | v_i$ is the end of the $i-$th sub-walk $W_{h,v_i}^{i}\}$. Thus, probability $pe(v, h)$ may be written as

follows,

$$\begin{aligned} pe(v,h) &= ph \cdot \left(1 - \mathbb{P}\left(\bigcap_{i=0}^{F-1} E_i\right)\right) \\ &= ph \cdot \left(1 - \mathbb{P}(E_0) \cdot \prod_{i=1}^{F-1} \frac{\mathbb{P}(\cap_{k=0}^{i} E_k)}{\mathbb{P}(\cap_{j=0}^{i-1} E_j)}\right) \\ &= ph \cdot \left(1 - (1 - \mathbb{1}_{\{v=h\}}) \cdot \prod_{i=1}^{F-1} \frac{\sum_{x\in\{1,\ldots,N\}} (\mathcal{M}_{(v)}^{i})_{h,x}}{\sum_{x\in\{1,\ldots,N\}} (\mathcal{M}_{(v)}^{i-1} \cdot \mathcal{M})_{h,x}}\right), \end{aligned}$$

where the second equality comes from repeated application of the definition of conditional probability. The third equality comes from the definition of probability as the ratio of favorable outcomes and total cases. For both probabilities we can take the same number of total cases as the number of $i-$walks, which cancel in the division and thus only the favorable cases survive. In the first probability, the favorable cases are the number of $i-$walks from h never getting v, which are estimated by the $i-$th power of matrix $\mathcal{M}_{(v)}$, the adjacency matrix of D where all ingoing and outgoing arcs of v are removed. Favorable cases at denominator is just the number of $i-$walks from h never touching v except maybe in their last step. If there are no $i-$walks from h in D, for consistency we define this ratio of probabilities to be 1. Now, when there are $H \geq 1$ hunters $\{h_i\}_{i=1}^{H}$, the *Inclusion-Exclusion Principle* allows us to generalize the probability of eating $pe(v, h_i)$ from hunter h_i, to the probability of getting food from some hunter $pe(v)$,

$$pe(v) = \sum_{k=1}^{H} (-1)^{k-1} \sum_{I\subseteq\{1,\ldots,H\},|I|=k} \prod_{i\in I} pe(v, h_i). \tag{D.1}$$

This expression is valid since we have assumed that generation of food by every hunter correspond to independent events.

D.2 Structure of simulation and sampling of model variables

The methodology we use to analyze our model is to sample a representative set of values for the model variables, obtain for each combination of model variables the respective set of optima via evolutionary algorithms, and then describe these optima with the pipeline of analysis. We use as a study case a population size of $N = 12$, since it is one of the smallest numbers allowing for exploring combinations of number of groups ng and group sizes gs, where both may hold that $ng, gs \geq 3$ and simultaneously one of them may be strictly greater than 3. These magnitudes are enough to represent the network of food transfers of a typical cluster of households, each formed by 3 to 4 families, and each family usually comprised by an adult couple and their dependent children [89]. It is clear that using larger population sizes may produce a wider diversity of structures that may qualify as optimal networks for the model, and that it is not straightforward to define a similarity notion between classes of

optimal networks from distinct population sizes in order to assess the scalability of results obtained for a particular network size. However, the empirical fact that larger camps or hunter-gatherer communities are comprised by larger numbers of clusters rather than by larger clusters [89], is suggestive that there may be some kind of spatial constraints in play, such as food productivity by land unit or land limits to household habitability, that shape larger camps as an aggregation of typical-size-clusters. Exploring these spatial constraints is out of the scope of our work, and our approach is to explore the possible network organization with the typical size of a cluster of households.

Regarding a possible parameterization of a limited number of parameters, we have only fixed the RV function by fixing the values of k and n -the critical time period and the life span-. Since distinct RV curves have a similar behavior in general terms, see Figure 4.2, we estimate that fixing this function does not impose a hard constraint to the analysis, because it only affects the ranges of ph where distinct regimes are observed. We do not have any *a priori* hypothesis on a potential dependence of model variables, and nor know whether it may simplify the analysis. The previous evolutionary model CURP [227], which models the emergence of food-sharing in function of resource pressure and has two variables similar in functionality to ph and F (*prob-resource* to ph, and *min-energy* to F), does not give evidence that these two variables could be easily simplified into one variable. However, on the Annex D.6 there is an approximation for the mean probability of eating for WEF optima, $\overline{pe}(v) \sim \frac{ph}{N} \cdot (F \cdot nh)$, which we conjecture that may serve to establish some invariant on WEF networks with different sizes and similar values of this mean probability. Based on this expression, we have chosen the approach of separately sampling, on the one hand, ph, and on the other hand, F and nh. The intuition to sample the first variable will come from examining the RV function which we will work with. In the case of the last two variables, it comes from the fact that the product $(F \cdot nh)$ represents the maximum of times some food is ever shared in the network. Hence, whether this amount is enough to feed the whole network, that is, whether $(F \cdot nh) < N$, is an important variable for network organization of WEF optima.

The intuition to sample the ph values is the following. The reduction of variability function, which we work with has a maximum value at approximately $pe^\star = 0.0899$. Since non-hunter nodes usually eat with a probability smaller than ph, the simpler way of minimizing RV for values of ph smaller than $pe^\star$, is that non-hunters just do not receive food and are isolated nodes. This trend is progressively reversed if $ph > pe^\star$, since there is a cost in increased RV mean by leaving behind non-hunters, which is accentuated for greater values of ph where RV reaches its minimum values. Following this intuition, we have chosen the values $\{0.02, 0.08, 0.15, 0.3, 0.6\}$ of ph by choosing two smaller, two greater than $pe^\star$, and one at the RV flat region. Now, for the other two variables, we have chosen a set of 10 value pairs (F, nh), which is balanced with respect to condition $F \cdot nh < N = 12$. Since there are more combinatorial network-walks possibilities when $F > 2$ and this makes these cases more interesting, the specific value of F constrains nh to only certain possible values in the case $F \cdot nh < N$. This constraint for nh is not present if $F \cdot nh \geq N$. On the other hand, we conjecture that for general networks of size N, the greatest variations in the probability of eating are produced by the range $F \in \{2, \ldots, N/2\}$, and that larger values of F produce smaller variations. Guided by these assumptions, we have sampled the following 10 pairs of

variables (F, nh).

$$\{(12,1),(4,3),(3,5),(3,7),(4,9),(4,1),(4,2),(5,2),(3,3),(2,5)\}$$

We have taken each possible combination of the 50 (ph, F, nh) values as initial condition for the evolutionary network minimization of criteria combination.

D.3 Sizes of tables of networks and features

The single optima datasets, for RV and WEF, both were appropriately processed by tSNE clustering method, by gathering the data of all conditions of model variables in a single table. The table sizes are the following. The number of features are explained in the section describing them.

Table D.1: **Dimensions of single criteria optimal networks datasets.**

Optima type	**Number of networks**	**Number of features**
RV	6047	10
WEF	18686	12

In the case of Pareto optimal networks, however, the set of networks is too diverse to be properly handled in a single table. For this reason, the networks data was split according to the value of ph and the condition whether $F \cdot nh \leq N$. See the paragraph *Summary of distinct types of WEF optima* from Section 4.3.1, for the result justifying the use of this last condition. Having said that, below are the sizes of tables used to cluster the datasets of Pareto optimal networks. All these tables have the same feature set used in the clustering of RV data. All datasets and code used to generate them may be consulted in the repository provided for this purpose [233].

Table D.2: **Dimensions of multicriteria optimal networks datasets.**

ph	$F \cdot nh \leq N$	**Number of networks**
0.02	True	54503
0.02	False	13642
0.08	True	46522
0.08	False	14703
0.15	True	48910
0.15	False	1852
0.3	True	11082
0.3	False	611
0.6	True	3186
0.6	False	109

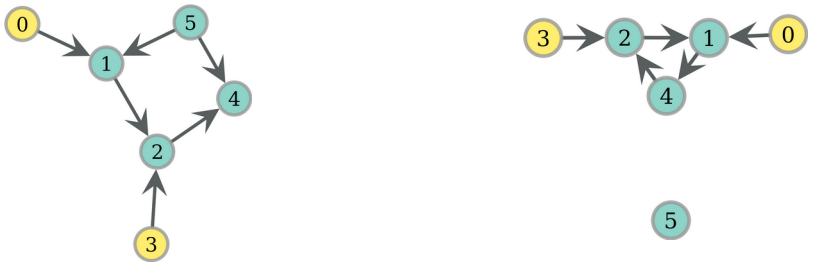


Figure D.1: **Two non-isomorphic networks with the same cost for $F = 3$.** Networks D_1 (left) and D_2 (right). Hunters $\{0, 3\}$ are filled in yellow.

D.4 Domain of optimization model and evolutionary algorithm implementation details

As an example of the situation of parts of a network that do not contribute to network cost, consider the networks of size 6 in Figure D.1 with node hunters $\{0, 3\}$. Networks D_1 and D_2 are clearly non-isomorphic. But if we work with $F = 3$, the probabilities of eating only depend on the walks of length 2 starting from the hunters, which in D_1 are $\{(0, 1), (1, 2)\}$, $\{(3, 2), (2, 4)\}$, while in D_2 these are $\{(3, 2), (2, 1)\}$, and $\{(0, 1), (1, 4)\}$. These walks produce the same probabilities of eating for every node under the map $\pi : V_1 \to V_2$, $\pi(0) = 3$, $\pi(3) = 0$, $\pi(1) = 2$, $\pi(2) = 1$, $\pi(4) = 4$ and $\pi(5) = 5$, and in this way, the two networks have exactly the same value for the two criteria. Since the probabilities of eating depend on the reachability from hunters by $(F-1)-$walks, there will be a huge number of non-isomorphic networks that have the same objective value, due to subgraphs that are not reachable via $(F-1)-$walks starting from hunter nodes, as we have seen in this example. We remedy this by focusing on those regions of a network which indeed influence the probabilities of eating. Specifically, we will work with the following notion of isomorphism.

Definition 6 (Isomorphism with hunters) Let D, G two digraphs of N nodes, and two respective sets of hunters $H(D) \subseteq V(D)$, and $H(G) \subseteq V(G)$, of equal size, $|H(G)| = |H(D)|$. Let $F \in \mathbb{N}$ with $F \geq 2$. We say that D and G are $F-$*isomorphic with hunters*, if there is a bijection $\pi : V(D) \to V(G)$, satisfying the following:

- The set of hunters is an invariant under the bijection, that is, $\pi(H(D)) = H(G)$.
- For every $u \in V(D)$ at distance at most $(F-2)$ to some $h \in H(D)$, it holds that $(u, v) \in E(D)$ if and only if $(\pi(u), \pi(v)) \in E(G)$.

It is not difficult to see that the isomorphism with hunters is an equivalence relation in the set of directed graphs (ie, a binary reflexive, symmetric and transitive relation), and hence the set of digraphs can be partitioned into equivalence classes of isomorphic networks having the same costs, even though they may have distinct number of arcs, as seen in the last example. Given these elements, we claim that the *domain set for solution points* of our model corresponds to the set of representative digraphs of the isomorphism equivalence classes. This definition aims to discard portions of the digraph that do not contribute to the objective value.

We next review some implementation decisions we have taken in order to execute the evolutionary algorithms more efficiently, including the mutation procedure which leverages the definition of the domain of the problem we just reviewed. The individuals in the evolutionary algorithm, or in other words, the networks, are equipped with a data-structure that stores one vector for each hunter, that keeps the updated distance of each network node from the respective hunter. This data structure enables for implementing feasible mutations as those random arc additions or deletions that produce networks belonging to the set of feasible solutions, which reduces the exploration space in comparison to performing arbitrary random arc changes. Distance updates are made by running BFS-like (*breadth-first-search*) algorithms from the respective hunter when a relevant arc addition or deletion has been produced. The reader interested in consulting the specifics of the evolutionary algorithms used may review the code [233] written to perform the simulations.

Other important optimizations are those relative to the data-structure that stores the optima candidates of the optimization, which we will refer to as the *archive*. At each generation, when a population is reviewed to update the archive, the population is filtered, and only those individuals passing the filter are reviewed for possible addition to the archive. In the case of single objective optimization, the filter is being a local optima for the respective function. In the multi objective case, the offspring population is processed by NonDominatedSort, an efficient algorithm that performs a ranking according to Pareto domination [101], thus only the set of individuals that most likely belong to the Pareto optimal set are tested to be added to the archive. Now, it is straightforward to maintain in the archive the updated set of non-dominated individuals processed. However, there is no direct way to determine whether an individual belongs to the Pareto optimal set. Therefore, we have implemented an adaptive heuristic to determine when to stop the generation process such that the archive likely stores a high proportion of Pareto optimal networks. This heuristic is rooted on our empirical observation on exhaustive optima search on networks of small size ($N = 5$), that the number of dominated networks that are erased from the archive at each iteration, becomes negligible with respect to the size of the archive, when the latter converges to a high proportion of Pareto optimal networks. Thus, our heuristic computes the ratio of the standard deviation of erased networks in the last and antepenultimate iteration windows of size $w = 10$, and the current size of the archive. The evolution is stopped if these 2 ratios are smaller than 0.005. On the other hand, single-objective minimization used 2000 generations.

D.5 Choice of tSNE hyperparameters and heuristics to set OPTICS hyperparameters

The choice of parameters is critical to accomplish a good visualization with tSNE. Since we work with relatively large datasets, we have used the following recommendations [167]: perplexity as 1% of the dataset size n -or the number of networks-, PCA initialization, high learning rate of n/EE, where EE is the early exaggeration, which is chosen usually as 4 or a small number to speed the convergence to solution [28]. We found that this parameter instantiation yielded robust results for distinct random seeds. We used the sci-kit learn implementation of tSNE [224] together with the extension intelex (https://intel.github.io/scikit-learn-intelex/)

to accelerate computations of this library.

Now, a cluster in OPTICS is the set of all *density-reachable* objects, or objects connected by regions of similar density, from an arbitrary *core object*, which is a point with at least $MinPts$ points in a neighborhood of radius ε named as the *reachability-distance*. OPTICS creates an ordering of the database, which enables producing cluster memberships in linear time for any given $\varepsilon \geq 0$. This parameter determines that objects with a reachability-distance greater than ε, to their closest density-reachable core object, are considered to be noise. We have chosen $MinPts$ to be the 3% of the dataset size, a relatively large magnitude which reduces the number of core points, and hence, the number of clusters. In order to choose ε, we have implemented the following iterative heuristic, which looks for an epsilon producing a minimun of intra-cluster dispersion in features, while keeping low: the size of noise, and its heterogeneity. At each step, pick the greatest reachability distance in the dataset, compute ε as the average of this distance and the greatest distance of last iteration, and obtain respective clustering with this ε. Compute the average, over clustering classes, of the sum of intra-class feature standard deviation, for 3 cases: all the labels ($\bar{\sigma}_a$), all the labels except noise ($\bar{\sigma}_{a'}$) and only noise ($\bar{\sigma}_n$). Store the current ε if minimizes $\bar{\sigma}_a$ and $\bar{\sigma}_n < rst \cdot \bar{\sigma}_{a'}$, where rst stands for *ratio of standard deviation*. Remove object with the greatest distance and go to next iteration, up to removing a proportion mnp -for *maximum noise permitted*- of the original dataset.

Now, the values of mnp, rst are usually around 0.02 and 1.5. These are the values used for single optima datasets. However, in multicriteria optima datasets, due to greater diversity, these values were checked and eventually reassigned by inspecting that the visual aspect of OPTICS clusters of tSNE map, was not too raw, with a few overly large clusters. This is usually accomplished with a value of $\bar{\sigma}_a$ approximately between 25 and 35. The parameters used for multicriteria optima are in the table below. All code implementing this heuristics may be consulted in the repository provided for this purpose [233].

Table D.3: **Parameters for OPTICS heuristics on multicriteria optima.**

ph	$F \cdot nh \leq N$	mnp	rst
0.02	True	0.02	4
0.02	False	0.02	3
0.08	True	0.02	3
0.08	False	0.03	4.5
0.15	True	0.03	3
0.15	False	0.07	4.5
0.3	True	0.02	4
0.3	False	0.11	3
0.6	True	0.03	7.5

D.6 Analytical argument for the inclusion of additional features in the construction of decision trees

The quantity $F \cdot nh$, which intuitively represents the maximum of times some food is ever shared in the network, approximately determines the mean of the probability of eating for WEF optima, as we review next. If we start from (D.1) to write the mean probability of eating for v non-hunters, and since $pe(v, h_i) \in (0, 1), \forall i$, we can take a first order approximation by neglecting products of $pe(v, h_i)$ for $k > 1$, leading to

$$\bar{pe(v_{\text{non-h}})} \sim \sum_{i=1}^{H} pe(v, h_i) = nh \cdot pe(v, h) \sim nh \cdot \frac{(F-1) \cdot ph}{(N - nh)}, \tag{D.2}$$

where the last approximation comes from the equidistribution, on WEF optima, of $(F-1)$ events of food sharing with probability ph, over a total of $(N - nh)$ non-hunters. Thus, an expression for the mean probability of eating may be written by splitting the mean into hunters and non-hunters, and replacing by (D.2)

$$\bar{pe(v)} = \frac{nh \cdot ph}{N} + \frac{(N - nh)}{N} \cdot \bar{pe(v_{\text{non-h}})} \tag{D.3}$$

$$\sim \frac{ph}{N} \cdot (F \cdot nh). \tag{D.4}$$

Equation (D.3) justify our decision to add some model variables to the feature set employed to train the classification trees.

D.7 Implementation details of decision trees

We implemented the model of decision trees we have used in Python, starting from an open-source implementation of model-trees available on the following site.

`https://github.com/cerlymarco/linear-tree`

We measure syntactic stability of a tree following the approach by Dunne et al (2002) [88] where stability is defined as the average of some similarity measure over the set of distinct pairs of trees belonging to a set of M tree instantiations. The DT similarity measure employed is minus the tree-edit distance of the pair of trees, computed by the APTED algorithm [223] and implemented by an open-source project that may be found in the site below.

`https://pypi.org/project/apted/`

In order to avoid a large number of similar trees with similar values of accuracy and stability, we have grouped those trees that have the same structure and split variables into the same tree

class, on which the average of stability and accuracy is computed. We have used $M = 15$ tree instantiations to compute the accuracy and stability pair for each set of DT hyperparameters, where each instantiation comes from a distinct seed of bootstrap sampling. The set of $HP = 1000$ DT hyperparameters is obtained via a random uniform search [32] over the entire set of hyperparameters, which is described in the following table. These ranges were chosen by empirically checking whether they give rise to high accuracy values, usually superior to 0.7. Floating points for number of samples in this table are interpreted as proportions of the dataset. All code implementing these classification trees may be consulted in the repository provided for this purpose [233].

Table D.4: **Ranges of decision tree hyperparameters.**

Hyperparameter	**Lower limit**	**Upper limit**
minimum of samples to perform a split	$3 * 10^{-5}$	$3 * 10^{-5} + 10^{-3}$
minimum of samples per leaf	$3 * 10^{-5}$	$3 * 10^{-5} + 10^{-3}$
maximum depth	3	8
minimum impurity decrease	0	0.01

Finally, since we work with accuracy and stability sample averages that aim to approximate population parameters, we used the following procedure to determine statistical Pareto-dominance. These sample averages follow approximately a normal distribution, which allows for performing the usual $t-$tests for mean comparison. In order to determine if the pair $\vec{u} = (a_u, s_u)$ Pareto-dominates $\vec{v} = (a_v, s_v)$ (as a maximization), we compute a $2-$sample $1-$tail mean $t-$test for accuracy and stability. If the null hypothesis is rejected, that is, we cannot reject that $a_u < a_v$ or $s_u < s_v$, we know that $\vec{u}$ does not Pareto-dominate $\vec{v}$. Otherwise, compute a $2-$sample $2-$tail mean $t-$test for accuracy and stability. If the null is rejected, that is, we cannot reject that $a_u \neq a_v$ or $s_u \neq s_v$, we conclude that $\vec{u}$ Pareto-dominates $\vec{v}$.

D.8 Decision trees for Pareto optimal networks

Networks on Figure 4.11 correspond to the most central networks on the following leaves. Figure 4.11 (a) from leaf 22 on tree at Figure D.2. Figure 4.11 (b) from leaf 14 on tree at Figure D.3. Figure 4.11 (c) from leaf 19 on tree at Figure D.4. The interested reader in examining these or other trees describing sets of optimal networks, may review the documentation for running the code from the repository provided for this purpose [233].

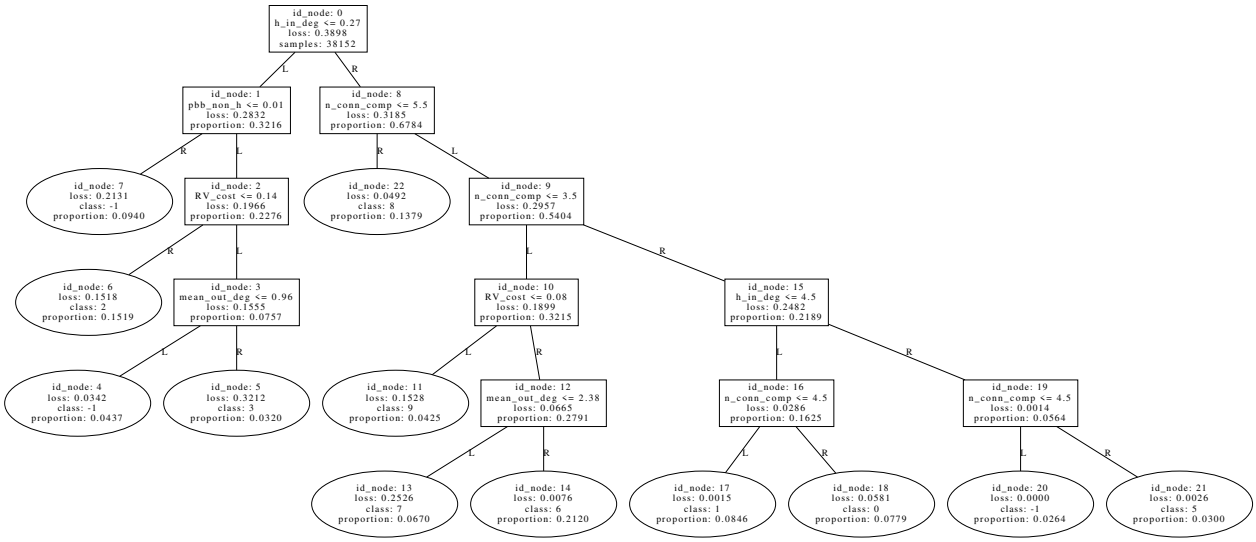


Figure D.2: **An efficient tree (accuracy = 0.747) to discriminate the clustering labels of PF optima on** $ph = 0.02$**.** Statistics in the tree are computed on the training set, while average accuracy is computed in the test set. See Annex D.3 for the sizes of the datasets used, and Paragraph *Description of clusters by classification trees* (Section 4.2.5) for the general procedure of tree construction. See the first paragraph from Section 4.3.1 for an explanation of the variables displayed in tree nodes.

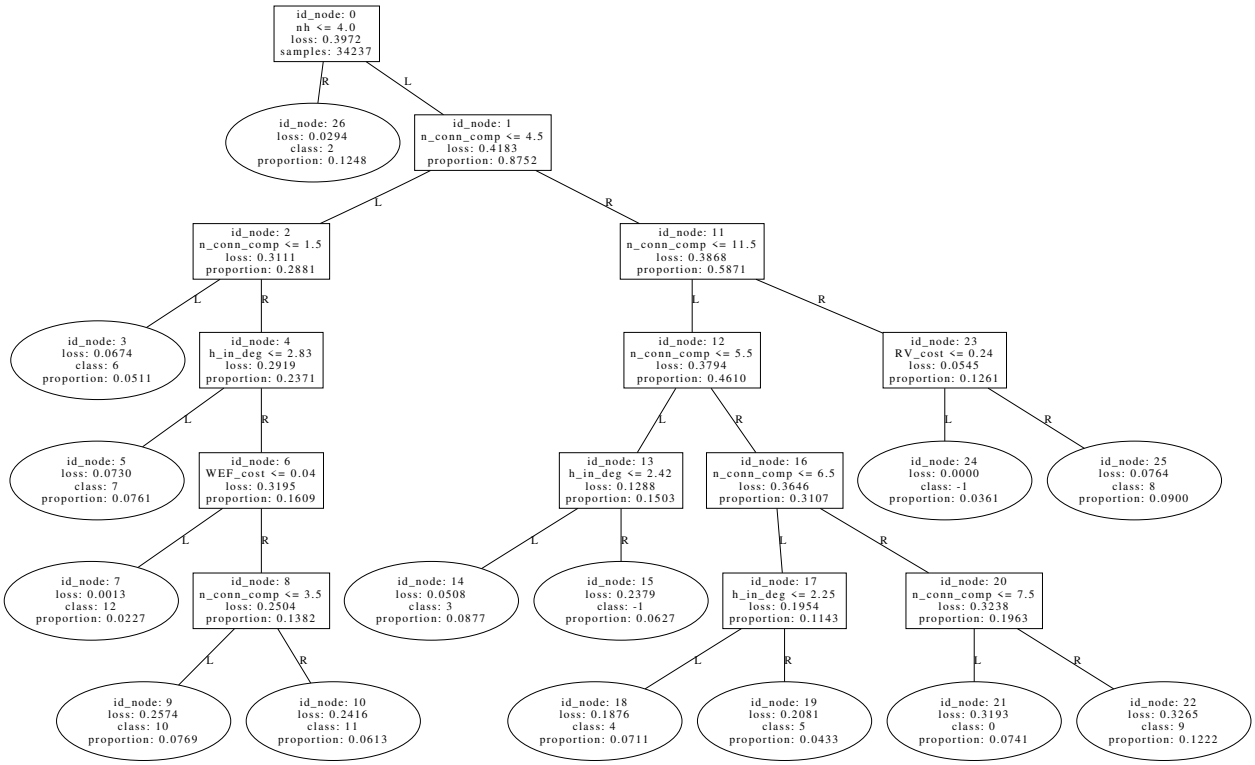


Figure D.3: **An efficient tree (accuracy = 0.734) to discriminate the clustering labels of PF optima on** $ph = 0.15$**.** Statistics in the tree are computed on the training set, while average accuracy is computed in the test set. See Annex D.3 for the sizes of the datasets used, and Paragraph *Description of clusters by classification trees* (Section 4.2.5) for the general procedure of tree construction. See the first paragraph from Section 4.3.1 for an explanation of the variables displayed in tree nodes.

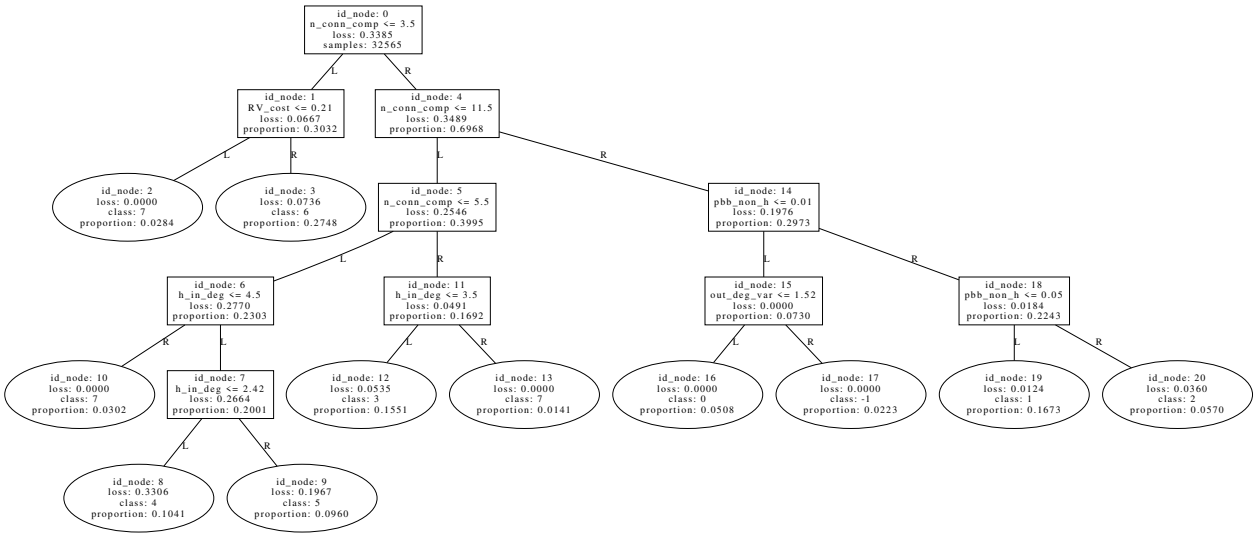


Figure D.4: **An efficient tree (accuracy = 0.845) to discriminate the clustering labels of PF optima on** $ph = 0.08$**.** Statistics in the tree are computed on the training set, while average accuracy is computed in the test set. See Annex D.3 for the sizes of the datasets used, and Paragraph *Description of clusters by classification trees* (Section 4.2.5) for the general procedure of tree construction. See the first paragraph from Section 4.3.1 for an explanation of the variables displayed in tree nodes.

Annex E

Are food-sharing networks scale-free?

Abstract

This annex of the research shown in Chapter 4, displays a series of analysis made later to the paper submission, with the goal of giving an initial answer to the following question. Are the optimal networks from this food-sharing model able to recover the scale-free property? It is an important question considering that this property, together with the community structure, are the main themes of this thesis, and this model was designed with the goal of studying only a hypothesis for community structure in a specific type of social networks. We have used a simple method to fit the parameters of the discrete power-law distribution to the in-degree and out-degree distributions of the optimal networks of the food-sharing model, and despite all the limitations because of the network size, the analysis suggests that some WEF and PF optimal networks, for larger sizes, may exhibit scale-free patterns. All the code and datasets generated are available online, together with another set of analyses performed, with negative results.

E.1 Introduction

The two main network patterns addressed in this thesis are, the power-law distribution, and the community structure. Since the food-sharing model studied in Chapter 4 is a hypothetical mechanism for the community structure, it is natural to ask whether this model also exhibits a degree distribution with a power-law or similar distribution. This point finds support in the fact that some empirical food-sharing networks have been found to present fatter in-degree tails than randomly formed groups of the same size [11]. In this Annex we make an initial exploration of this issue by analysing the optimal networks from our food-sharing model. The results, in spite of being tentative, offer a reasonable perspective about which of the optimal networks from our food-sharing model, may also qualify, for eventually larger network sizes, as scale-free networks.

E.2 Methods

Each optimal network found by the evolutionary optimization of the food-sharing model is analyzed with the following procedure. Either the in-degree and out-degree distribution of each network, are fitted by the discrete power-law distribution. This distribution is defined for a discrete set of integer values, and has an analytic expression for its density which

is analog to the one reviewed in Section 2.2.3 for continuous power-law distributions. Its normalizing constant is given by the reciprocal of the Hurwitz zeta function [67],

$$\mathbb{P}(X = x) = x^{-\alpha}/\zeta(\alpha, x_{\min}), \quad \zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty}(n + x_{\min})^{-\alpha}.$$

For this distribution, there is not an explicit expression for the MLE estimator of the exponent α, but instead this estimator can be found by numerical maximization of the logarithm of the likelihood [67]. The parameter $x_{\min}$ may be estimated with the same method reviewed in Section 2.3.3, and the p-value introduced in Section B.2 may be used as a measure of goodness of the power-law fit. This p-value corresponds to the probability that any random model gets a poorer fit, in terms of the distance to the empirical distribution, than that given by the specific estimated $\hat{x}_{\min}$ and $\hat{\alpha}(\hat{x}_{\min})$ values. This p-value is the main index we use in this section to assess whether the estimated parameters offer an acceptable fit, corresponding to a p-value nearer to 1, or a bad fit, corresponding to a p-value close to 0.

We used the R package *poweRlaw* [111] to obtain these parameter estimates, where both $x_{\min}$ and its p-value are approximated by Monte Carlo bootstrap computations. We have used the same heuristic constraint applied in Section B.1 of limiting the search space of $x_{\min}$ to $(1, q_{20})$, where q_{20} is the percentile 20 of the respective (in- or out-)degree distribution.

E.3 Limitations

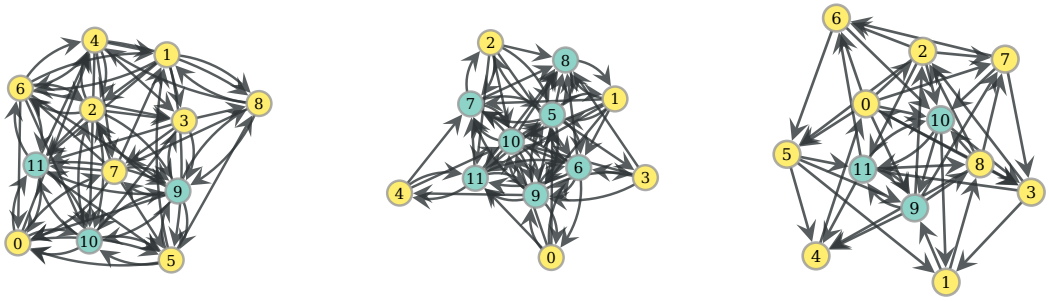
The conclusions of this analysis should be taken as tentative, for a number of reasons. The small size of the networks ($N = 12$) limits either the size of the samples, but also, since the networks are simple, the possible values taken by the degree values, an aspect that is important considering the properties of the power-law distribution. The fitting procedure we use presents a consistency that has been theoretically proven only for some simple models [87]. Other methods for estimating the power-law exponent have better consistency properties [281], but under the asymptotic limit of the sample size. However, as the following section shows, the results obtained are reasonable as a first approximation.

E.4 Results

Regarding WEF optimal networks, we use the same tree from Figure 4.3, to visualize which type of networks are more likely to be well fitted by a power-law degree distribution. Table E.1, show the only tree leaves where the variables p-value of out-degree (p_{out}), exponent of out-degree (α_{out}), p-value of in-degree (p_{in}), and exponent of in-degree (α_{in}), get positive values on the respective percentile statistics. From the leaf numbers, it is apparent that the WEF networks whose degree distribution is better approximated by a power-law, are some of those where $(F \cdot nh) > N$, that is, networks where there is enough supply of food for every node, as such in Figs. E.1a, E.1b.

Table E.1: **General statistics of power-law fit on WEF optima.** Leaves are from the tree of Figure 4.3.

Feature	Leaf	Perc. 0.5	Perc. 25	Median	Perc. 75	Perc. 99.5
p_{out}	13	0.0	0.0	0.0	0.53	0.65
p_{out}	15	0.0	0.0	0.0	0.53	0.65
α_{out}	13	0.0	0.0	0.0	2.81	4.52
α_{out}	15	0.0	0.0	0.0	2.36	3.69
p_{in}	13	0.0	0.0	0.0	0.53	0.87
p_{in}	15	0.0	0.0	0.0	0.48	0.56
α_{in}	13	0.0	0.0	0.0	4.35	9.0
α_{in}	15	0.0	0.0	0.0	2.41	3.43



(a) Leaf 13 (b) Leaf 15 (c) Leaf 4

Figure E.1: **Randomly sampled networks from the distinct tree leaves of single optima.** These networks are randomly sampled from the subset of networks of each leaf where the p-value of the out-degree distribution is greater than 0.5. Hunters are filled in yellow. Networks E.1a, E.1b are WEF optima from the tree from Figure 4.3. Network E.1c is a RV optimum from tree from Figure 4.8.

These leaves are characterized by having the greatest values of standard deviation of out-degree distribution among leaves, as can be seen in Table E.2, which may help explaining the fact these were the only leaves with positive p-values for the power-law fit. This is because the rather small values of the fitted out-degree exponents would be consistent with the theoretical behavior of scale-free networks described in Section 2.2.5, where it is explained that one of their key patterns is the divergence of the second moment when $\alpha \leq 3$. On the other hand, the differences between leaf 13 and 15 are, besides the number of hunters (9 in leaf 13, and 5 to 7 in leaf 15), in the distribution of the standard deviation of in-degree (Table E.2), which by the same argument given above, is consistent with the difference existent among these leaves in the exponents of in-degree distribution. In the case of RV optima, by making the same analysis and working with the tree of Figure 4.8, now leaf 4 is the only one with positive p-values. This leaf, as explained in Section 4.3.2, includes networks with patterns similar to WEF optima as such in Figure E.1c. The fact that several of these networks minimize in-degree variability is consistent with the exponent of in-degree distribution obtained in Table E.3.

Now, we review the PF optimal networks. In the tree shown in Figure D.2 for $ph = 0.02$ and $F \cdot nh \leq N$, the only leaves with positive p-values are the 11, 13, and 14. Representative

Table E.2: **General statistics of standard deviation of degree distributions on WEF optima.** Leaves are from the tree of Figure 4.3. Features described are the standard deviation of the out-degree distribution (σ_{out}), and of the in-degree distribution (σ_{in}). The symbol **wh** means that the statistics are obtained on the whole dataset, not restricted to any leaf.

Feature	Leaf	Perc. 0.5	Perc. 25	Median	Perc. 75	Perc. 99.5
σ_{out}	13	1.07	1.5	1.66	1.89	2.62
σ_{out}	15	1.07	1.48	1.65	1.82	2.29
σ_{out}	**wh**	0.64	1.23	1.49	1.74	2.84
σ_{in}	13	0.62	0.96	1.04	1.14	1.44
σ_{in}	15	1.6	2.06	2.33	2.62	3.22
σ_{in}	**wh**	0.86	1.64	1.93	2.5	3.22

Table E.3: **General statistics of power-law fit on RV optima.** Leaves are from the tree of Figure 4.8.

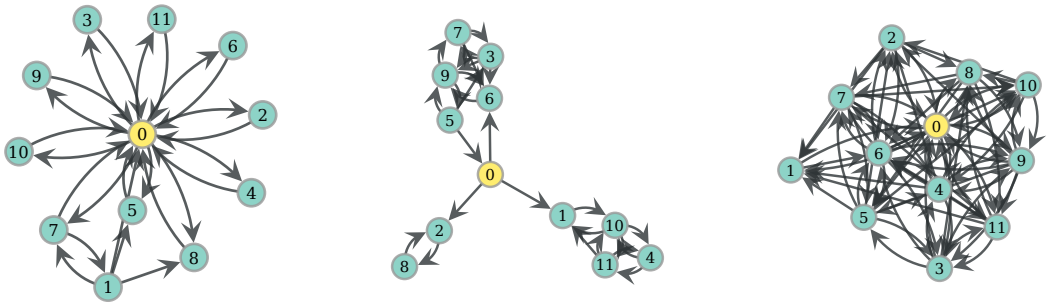
Feature	Leaf	Perc. 0.5	Perc. 25	Median	Perc. 75	Perc. 99.5
p_{out}	4	0.0	0.0	0.0	0.0	0.63
α_{out}	4	0.0	0.0	0.0	0.0	1.96
p_{in}	4	0.0	0.0	0.0	0.0	0.97
α_{in}	4	0.0	0.0	0.0	0.0	7.75

networks from these tree leaves are plotted in Figure E.2. The statistics of power-law fit are presented in Table E.4. In the first network, the power-law behavior may be intuitively explained by the differences in the out, and in-degree between the hunter at the center, and the peripheral non-hunters. The magnitudes of exponents for leaf 14 may be related to the fact that this leaf is the one minimizing the standard deviation of out-degree, see Table E.5. The case of networks in leaf 13 is difficult to explain, in part because these are outliers of the p-value distribution, see Table E.4. The networks in leaf 13 minimize the mean out-degree, at least in the median and upper percentiles, see Table E.5, which seems contradictory with the power-law exponents $\alpha < 2$ in this leaf. Since additionally the p-values on this leaf are not high, all these elements support the hypothesis that these p-values are likely a fitting artifact.

Networks with positive p-values in the trees where $ph = 0.08, 0.15$ (Figures D.4,D.3) are the same kind of networks from leaves 11 and 14 in Figure E.2. Multiobjective networks obtained for $ph = 0.6$ do not obtain positive p-values.

E.5 Conclusions

Despite all the limitations of our datasets because of the network size, the simple method we have used allows suggesting potential networks of our food-sharing model that, for larger sizes, may exhibit the scale-free patterns typical of the power-law distribution for $\alpha \leq 3$. These networks are mainly some WEF optima like those in Figure E.1b, and some PF optima like those in Figure E.2a. We do not have intuition regarding the correlation between p-values



(a) Leaf 11 (b) Leaf 13 (c) Leaf 14

Figure E.2: **Randomly sampled networks from the distinct tree leaves of multiobjective optima.** These networks are randomly sampled from the subset of networks of each leaf where the p-value of the out-degree distribution is greater than 0.5. Hunters are filled in yellow. Networks E.2a, E.2b and E.2c are, respectively, from tree nodes 11, 13 and 14 from the tree in Figure D.2.

Table E.4: **General statistics of power-law fit on PF optima.** Leaves are from the tree of Figure D.2.

Feature	Leaf	Perc. 0.5	Perc. 25	Median	Perc. 75	Perc. 99.5	Perc. 100
p_{out}	11	0.0	0.0	0.0	0.0	0.76	0.81
p_{out}	13	0.0	0.0	0.0	0.0	0.0	0.51
p_{out}	14	0.0	0.0	0.0	0.0	0.64	0.81
α_{out}	11	0.0	0.0	0.0	0.0	2.62	2.62
α_{out}	13	0.0	0.0	0.0	0.0	0.0	1.96
α_{out}	14	0.0	0.0	0.0	0.0	3.82	8.52
p_{in}	11	0.0	0.0	0.0	0.0	0.73	0.8
p_{in}	13	0.0	0.0	0.0	0.0	0.0	0.51
p_{in}	14	0.0	0.0	0.0	0.0	0.95	1.0
α_{in}	11	0.0	0.0	0.0	0.0	2.47	2.62
α_{in}	13	0.0	0.0	0.0	0.0	0.0	1.99
α_{in}	14	0.0	0.0	0.0	0.0	10.13	11.28

Table E.5: **General statistics of degree distributions on PF optima.** Leaves are from the tree of Figure D.2. Features described are the standard deviation of the out-degree distribution (σ_{out}), and the mean out-degree ($\bar{\text{deg}}_{\text{out}}$). The symbol **wh** means that the statistics are obtained on the whole dataset, not restricted to any leaf.

Feature	Leaf	Perc. 0.5	Perc. 25	Median	Perc. 75	Perc. 99.5	Perc. 100
σ_{out}	14	0.76	1.83	2.17	2.6	3.56	3.79
σ_{out}	**wh**	0.76	2.06	2.47	2.81	3.64	4.15
$\bar{\text{deg}}_{\text{out}}$	13	1.5	1.75	1.92	2.08	2.33	2.33
$\bar{\text{deg}}_{\text{out}}$	**wh**	0.25	1.67	2.08	2.92	6.5	6.75

of in-degrees and out-degrees. They may be related to finite-size effects, but testing this hypothesis would require contrasting with larger networks.

E.6 Other analyses

We have analysed these datasets with at least two other methods, with negative results. The first is the use, instead of the degree distribution, of the degree-degree distance [298] of the underlying undirected graph, which is an intrinsic measure defined for each network arc, and that has been suggested to present a more prevalent and smooth power-law than the degree distribution. The p-values of fitting a power-law over the distribution of this measure gives almost no differences between tree leaves among the distinct type of optima, which suggests that the fitting does not accomplish to associate any network pattern to the power-law behavior, and may be considered as noise. The second method we have used is that of computing the estimators of the tail exponent of regularly varying distributions [281], which, rather than delivering a p-value, are distinct estimators of the exponent of the tail of the power-law distribution based on order statistics, that present statistical consistency for the limit of asymptotic sample sizes. We computed the three estimators used by Voitalov et al (2019) [281] for either the in-degree and out-degree distributions, but we did not obtain a consistent behavior among them, probably due to the small samples (network sizes) used. The datasets and methods produced for these analyses are available, together with the main analyses, in the link shared below.

E.7 Data availability

All the code, datasets, as well as the documentation for running the methods to get the results presented in this section, are available in the following link.

`https://github.com/PanchoTonho/Scale_free_food_sharing`