



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

UNA ESTRATEGIA PARA REALIZAR EL PROCESAMIENTO SEMI-AUTOMÁTICO DE
ASPECTOS SOCIALES DE UNA COMUNIDAD VIRTUAL

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

FELIPE IGNACIO AGUILERA VALENZUELA

PROFESOR GUÍA:

SERGIO F. OCHOA DE LORENZI

PROFESOR CO-GUÍA:

SEBASTIÁN RÍOS PÉREZ

MIEMBROS DE LA COMISIÓN:

CLAUDIO GUTIÉRREZ GALLARDO

AIDAN HOGAN

VALERIA HERSKOVIC MAIDA

SANTIAGO DE CHILE

2016

Resumen

La masificación de servicios basados en Internet ha permitido la proliferación de nuevas estructuras sociales, como por ejemplo las comunidades virtuales o las redes sociales en línea, donde los miembros interactúan a través de tecnología mediada por computador. A pesar del conocimiento existente acerca de qué aspectos son importantes para diagnosticar el estado de una comunidad virtual, poca atención se ha puesto en cómo medir muchos de estos aspectos, a pesar que usualmente la información requerida para calcularlos está disponible en las plataformas de apoyo a estas comunidades. Esto ocasiona que los administradores no cuenten con información suficiente para diagnosticar o saber cuándo y cómo intervenir la comunidad para lograr un cierto objetivo.

Para ayudar a paliar dicho desafío, este trabajo de tesis propone una estrategia para realizar el procesamiento semi-automático de ciertos aspectos sociales, específicos de comunidades virtuales que utilizan un sistema de software como mediador de las interacciones entre sus miembros. A diferencia de otros trabajos del área, esta tesis se enfoca en el análisis de la estructura social que emerge de la comunidad, más que analizar los aspectos que motivan la participación de los miembros de la misma. Por lo anterior, los aspectos a considerar son aquellos relacionados con el concepto de sociabilidad: *personas*, *propósito* y *políticas*. Para lograr el objetivo antes mencionado, se proponen estrategias para capturar, procesar y representar, a través de indicadores, la información generada por una comunidad virtual.

En base a la literatura existente, se seleccionó un conjunto de indicadores propuestos en la literatura del área para cada aspecto social en estudio, entre los cuales se encuentran: niveles de participación, identidad, cumplimiento del propósito, claridad y niveles de moderación. Para cada uno de éstos se definió un proceso de evaluación, el cual incluye el uso de una o varias técnicas, entre las cuales se encuentran las siguientes: uso de encuestas, modelos de tópicos, y técnicas de centralidad en redes sociales, entre otras.

Adicionalmente, para poder hacer extensivo el presente trabajo al estudio de otras comunidades virtuales, se definió un modelo conceptual común para representar la información de las estructuras sociales en estudio. A partir de dicho modelo se realiza el procesamiento de los aspectos e indicadores mencionados anteriormente.

Finalmente se aplicó la estrategia de procesamiento en dos comunidades virtuales ya existentes. Los resultados obtenidos fueron evaluados por sus administradores, quienes indicaron que dichos resultados les permitieron conocer información acerca del funcionamiento y estructura de la comunidad virtual, que no poseían antes de este estudio. Con todo lo anterior, se espera que este trabajo ayude a desarrollar iniciativas futuras que sigan con este enfoque y faciliten el proceso de diagnóstico del estado de una comunidad virtual.

A veces los logros opacan el camino recorrido y nos hace olvidar que lo más importante son aquellas personas que nos ayudaron a recorrerlo, incluso en los momentos más difíciles.

Dedicado a mi familia, a mis padres, y especialmente a Raquel y Víctor.

Agradecimientos

Quiero aprovechar este espacio para agradecer a todos los colegas y amigos que contribuyeron de alguna forma en la elaboración de este trabajo. Quiero agradecer a José Ignacio Santa Cruz, Rafael Castillo y Luis San Martín, quienes colaboraron con toda su experiencia como administradores de las comunidades que fueron parte de este trabajo. También quiero agradecer a Sebastián Ríos y Sergio Ochoa, quienes me apoyaron y guiaron en todo momento en este proceso.

Tabla de Contenido

Capítulo 1 – Introducción.....	1
1.1 Descripción del Problema a Abordar.....	3
1.2 Objetivos de la Tesis	4
1.3 Metodología Utilizada	4
1.4 Contribución de la Tesis.....	5
1.5 Estructura de la Tesis.....	6
Capítulo 2: Marco Teórico y Trabajos Relacionados.....	7
2.1. Estructuras Sociales.....	7
2.1.1. Redes Sociales	7
2.1.2 Comunidades Virtuales.....	10
2.1.3 Comunidades Virtuales de Práctica.....	12
2.2 Estrategias de Evaluación de Comunidades Virtuales	15
2.2.1 El Enfoque Sicológico.....	15
2.2.2 El Enfoque Sociológico.....	16
2.3 Técnicas de Evaluación en Redes Sociales y Comunidades Virtuales.....	17
2.3.1 Análisis de Redes Sociales	17
2.3.2 Análisis de Comunidades Virtuales.....	19
2.4 Modelo de Conceptos Sociales.....	20
2.4.1 Friend of a Friend	20
2.4.2 Semantically-Interlinked Online Communities.....	21
2.4.3 SIOC-SNA-DM.....	22
Capítulo 3 – Estrategia de Evaluación Propuesta.....	26
3.1 Aspectos Sociales a Evaluar	26
3.2 Descripción de la Estrategia Propuesta	29
3.3 Visualización de los Resultados	30
Capítulo 4 – Los Aspectos Sociales en Detalle.....	33
4.1. Evaluación de Personas	33
4.1.1 Niveles de Participación	33
4.1.2 Proceso de Evaluación.....	36
4.1.3 Identidad	38
4.2 Evaluación del Propósito.....	40
4.2.1 Cumplimiento del Propósito.....	40
4.2.2 Claridad del Propósito	45

4.3 Evaluación de las Políticas	45
Capítulo 5 – Implementación de la Solución	48
5.1 Arquitectura General	48
5.2 Tecnologías Usadas en la Implementación	49
5.3 Despliegue de los Indicadores	50
Capítulo 6 – Estudio de Comunidad I.....	51
6.1. Análisis de la Comunidad.....	51
6.2 Proceso de Extracción de Datos	53
6.3 Evaluación de la Estrategia Propuesta – Personas.....	56
6.3.1 Preparación de Datos	56
6.3.2 Aplicación de las Técnicas de Evaluación	57
6.3.3 Evaluación de los Resultados	59
6.3.4 Presentación de los Resultados.....	64
6.3.5 Resultado de la Encuesta	66
6.4. Evaluación de la Estrategia Propuesta - Propósito	67
6.4.1 Preparación de los Datos	67
6.4.2 Aplicación de las Técnicas de Evaluación	68
6.4.3 Evaluación de los Resultados	69
6.4.4 Preparación de Resultados.....	74
6.4.5 Resultado de la Encuesta	75
6.5 Evaluación de la Estrategia Propuesta Políticas.....	75
6.5.1 Preparación de los Datos	75
6.5.2 Aplicación de Técnicas de Evaluación.....	76
6.5.3 Generación de Reporte con Resultados	77
6.6 Análisis de los Resultados Obtenidos.....	79
Capítulo 7 – Estudio de Comunidad II	82
7.1 Análisis de la Comunidad.....	82
7.2 Proceso de Extracción de Datos	83
7.3 Evaluación de la Estrategia Propuesta – Personas.....	84
7.3.1 Preparación de los Datos	84
7.3.2 Aplicación de las Técnicas de Evaluación	85
7.3.3 Evaluación de los Resultados	87
7.3.4 Generación de Reporte	88
7.4. Evaluación de la Estrategia Propuesta - Propósito	92

7.4.1 Preparación de los Datos	92
7.4.2 Aplicación de Técnicas de Evaluación	92
7.4.3 Evaluación de los Resultados	92
7.4.4 Generación de Reporte con Resultados	96
7.5 Evaluación de la Estrategia Propuesta Políticas	97
7.6 Análisis de los Resultados Obtenidos.....	97
Capítulo 8: Conclusiones y Trabajo a Futuro	99
Bibliografía.....	101
Anexo A. Publicaciones Relacionadas con la Tesis	108
Anexo B. Aspectos Sociales de las estructuras sociales en estudio, desde la perspectiva de diversos autores.....	109
Anexo C. Caracterización de las Redes Sociales Generadas para la Comunidad I.....	112
Anexo D. Tiempos de Procesamiento para la técnica PYTM para la Comunidad I.....	113

Índice de Imágenes

Figura 1. Representación de un sociograma.....	8
Figura 2. Tipos de comunidades virtuales.....	11
Figura 3. Evolución de una comunidad de práctica	13
Figura 4. Niveles de participación de los miembros de una comunidad	14
Figura 5. Modelo FOAF.....	21
Figura 6. Modelo SIOC simplificado.....	22
Figura 7. Modelo SIOC-SNA-DM simplificado.....	23
Figura 8. Interrelación de aspectos sociales que posee los sistemas de apoyo a comunidades virtuales.....	28
Figura 9. Esquema general de la estrategia propuesta.....	30
Figura 10. Visualización de la evolución temporal de un aspecto social en forma de gráfica.....	31
Figura 11. Utilización del modelo para evaluar una comunidad virtual.....	31
Figura 12. Proceso de Evaluación de Niveles de Participación.....	36
Figura 13. Formas de representar una red social a partir de interacciones entre los miembros ...	36
Figura 14. Representación del modelo LDA.....	41
Figura 15. Representación del modelo de tópicos Pitman-Yor.....	42
Figura 16. Proceso de Evaluación de Cumplimiento de Propósito.....	43
Figura 17. Proceso de Evaluación de Niveles de Moderación.....	46
Figura 18. Arquitectura del sistema de procesamiento y análisis.....	48
Figura 19. Pantalla con despliegue de indicadores.....	50
Figura 20. Actividad de la Comunidad I.....	52
Figura 21. Evolución de los niveles de participación de la Comunidad I.....	53
Figura 22. Diagrama de clases simplificado del Foro de Discusión de la Comunidad I.....	54
Figura 23. Evolución de las densidades para las redes sociales de la Comunidad I.....	56
Figura 24. Redes sociales para enero del año 2009 (2009-01) de la Comunidad I.....	57
Figura 25. Miembros que pertenecen a la periferia en la Comunidad I.....	59
Figura 26. Evolución de índice F para el periodo 2009-2010 de la Comunidad I, utilizando <i>degree</i> como medida de centralidad.....	60
Figura 27. Comparación del índice F de la Comunidad I, para las 6 técnicas aplicadas	60

Figura 28. Niveles de participación para la Comunidad I	61
Figura 29. Evolución de los niveles de participación de la Comunidad I	61
Figura 30. Evolución de los niveles de participación en forma porcentual de la Comunidad I... ..	62
Figura 31. Evolución de índice F para el periodo 2009-2010 de la Comunidad I	62
Figura 32. Comparación entre las diferentes técnicas SNA y conteo simple en la Comunidad I	63
Figura 33. Evolución temporal de niveles de participación para la Comunidad I.	64
Figura 34. Evolución de niveles de participación para la Comunidad I, en forma porcentual. ...	64
Figura 35. Resumen de los niveles de participación de la Comunidad I.....	66
Figura 36. Resumen de los tiempos de procesamiento de LDA y PYTM en la Comunidad I....	69
Figura 37. Comparación de técnicas LDA y PYTM en la Comunidad I	69
Figura 38. Estimación de la cantidad de tópicos para el año 2008 en la Comunidad I.....	70
Figura 39. Cantidad de tópicos para cada período para la Comunidad I.....	70
Figura 40. Cumplimiento de propósito para la Comunidad I.....	72
Figura 41. Comparación de resultado obtenido con PYTM y con conteo simple para la Comunidad I	73
Figura 42. Cruce entre resultados obtenidos con PYTM y conteo simple para la Comunidad I .	73
Figura 43. Resumen cumplimiento del propósito para la Comunidad I.....	74
Figura 44. Actividad de los administradores a través de la vida de la Comunidad I	76
Figura 45. Resumen de intervenciones realizadas por los administradores de la Comunidad I ..	77
Figura 46. Niveles de Moderación para la Comunidad I, considerando sus interacciones como moderación	77
Figura 47. Niveles de Moderación para la Comunidad I, sin considerar sus interacciones como moderación	78
Figura 48. Resumen de nivel de moderación para la Comunidad I.....	79
Figura 49. Resumen de los indicadores para la Comunidad I	80
Figura 50. Actividad de la Comunidad II.....	82
Figura 51. Miembros activos en la Comunidad II.....	83
Figura 52. Diagrama de clases simplificado del Foro de Discusión de la Comunidad II	83
Figura 53. Miembros pertenecientes a la periferia de la Comunidad II	86
Figura 54. Evolución de los niveles de participación de la Comunidad II.....	88

Figura 55. Comparación entre el mejor resultado obtenido con las técnicas SNA y conteo simple en la Comunidad II.	88
Figura 56. Evolución de los niveles de participación en forma porcentual en la Comunidad II..	89
Figura 57. Resumen miembros activos de la Comunidad II.	90
Figura 58. Resumen miembros pertenecientes al núcleo de la Comunidad II.	91
Figura 59. Tiempos de procesamiento para las técnicas PYTM y LDA en la Comunidad II.	92
Figura 60. Rendimiento considerando PYTM y LDA, para la Comunidad II, para el año 2013.	93
Figura 61. Cantidad de tópicos para cada período para la Comunidad II.	93
Figura 62. Estimación de la cantidad de tópicos para el año 2013 en la Comunidad II.	94
Figura 63. Cumplimiento de propósito para la Comunidad II.	96
Figura 64. Comparación de resultado obtenido con PYTM y con conteo simple para la Comunidad II.	96
Figura 65. Resumen cumplimiento del propósito para la Comunidad II.	97
Figura 66. Resumen de los indicadores para la Comunidad II.	98

Índice de Tablas

Tabla 1. Resumen de los aspectos sociales de las estructuras sociales en estudio, desde la perspectiva de diversos autores.	14
Tabla 2. Contextos e Interacciones para distintos sistemas de Software.....	49
Tabla 3. Participación de los miembros de la Comunidad I.....	52
Tabla 4. Correspondencia entre conceptos de la Comunidad I y el modelo social propuesto.	55
Tabla 5. Resumen de los datos obtenidos para la Comunidad I en el cálculo de los miembros pertenecientes al núcleo.....	58
Tabla 6. Rangos para determinar si niveles de participación son adecuados para la Comunidad I.	65
Tabla 7. Resultado de la Encuesta de Identidad para la Comunidad I.	67
Tabla 8. Resumen de los documentos para analizar de la Comunidad I.	68
Tabla 9. Ejemplos de tópicos con alta relación con la Comunidad I.....	71
Tabla 10. Ejemplos de tópicos sin relación con la Comunidad I.....	71
Tabla 11. Resumen de la clasificación de tópicos para la Comunidad I.	72
Tabla 12. Resultado de la Encuesta de Claridad del Propósito para la Comunidad I.....	75
Tabla 13. Resumen de las actividades de moderación realizada por los administradores de la Comunidad I.	75
Tabla 14. Resumen de la actividad de los administradores, para la Comunidad I.	76
Tabla 15. Correspondencia entre conceptos de la Comunidad II y el modelo social propuesto. .	84
Tabla 16. Resumen de las redes sociales generadas para la Comunidad II.....	85
Tabla 17. Miembros pertenecientes al núcleo en la Comunidad II, utilizando como técnica de evaluación <i>degree</i>	87
Tabla 18. Rangos para determinar si niveles de participación son adecuados para la Comunidad II.	89
Tabla 19. Resumen de los documentos para analizar de la Comunidad II.....	92
Tabla 20. Ejemplo de tópicos con alta relación con la Comunidad II.....	94
Tabla 21. Ejemplo de tópicos sin relación con la Comunidad II.....	95
Tabla 22. Resumen de la clasificación de tópicos para la Comunidad II.....	95

Capítulo 1 – Introducción

Einstein señalaba que en el siglo 20 han sido tres las bombas que han explotado con repercusiones mundiales: la demográfica, la atómica y la de telecomunicaciones [1]. Otros autores señalan que en este siglo ha ocurrido el “segundo diluvio”, o más bien, el diluvio de la información, refiriéndose al crecimiento exponencial, explosivo y caótico de la información, la cual ha seguido multiplicándose en forma sostenida a través del tiempo. En este escenario una nueva cultura ha surgido, en la cual el desarrollo de tecnologías digitales de información y comunicaciones ha sido vital, permitiendo que surjan nuevas condiciones y oportunidades para el desarrollo de los individuos y de la sociedad. El impacto provocado, tanto a nivel social como cultural, ha dado a la tecnología un rol preponderante dentro de la sociedad.

Para Levy [1] la tecnología, más que ser una entidad real que existe de manera aislada, es un punto de vista que enfatiza los componentes materiales y artificiales del fenómeno humano. Por tanto, es imposible separar lo humano de su entorno material. *La tecnología es producida por una cultura, y una sociedad es condicionada por esta tecnología* [1]. La tecnología provee acceso a ciertas posibilidades, cuyas repercusiones sociales podrían no existir sin su presencia.

Un ejemplo de lo planteado anteriormente es el continuo desarrollo que han tenido las tecnologías basadas en la Web, las cuales han posibilitado la masiva proliferación de entidades sociales basadas en dicha tecnología, como por ejemplo las comunidades virtuales o redes sociales en línea. Las personas dejaron de ser meros espectadores del contenido estático que predominaba la Web hace varios años atrás, tomando en la actualidad un rol mucho más participativo, opinando, colaborando mutuamente e incluso generando contenido en forma conjunta. La tecnología ha ofrecido a las personas un medio para poder interactuar con otras, a pesar de encontrarse ubicadas geográficamente en lugares diferentes, y les ha permitido crear un sentido de pertenencia e identidad, a pesar de no existir necesariamente un contacto cara-a-cara. La sociedad ha cambiado producto de lo anterior, provocando que el surgimiento de estructuras sociales no se encuentre condicionado a la presencia física de las personas en un mismo lugar, sino que la tecnología ha ofrecido un mecanismo para la formación de dichas estructuras en forma virtual. De esta manera, la tecnología ha cumplido un rol fundamental, siendo el mecanismo que permite mediar las interacciones que ocurren al interior de estas nuevas estructuras sociales, dado que a través de ellas sus miembros desarrollarán las características necesarias para su existencia.

Sin embargo, aplicar tecnología en este nuevo escenario no es una tarea sencilla. Como se mencionó anteriormente, ya no se trata de ofrecer sistemas que se comuniquen entre sí, ni de sistemas utilizados por usuarios en forma aislada, como lo sería un procesador de texto. Es necesario ofrecer el soporte adecuado a estas estructuras sociales de carácter virtual, de tal forma que sea posible mediar y facilitar la interacción entre sus miembros. De esta forma, surgen diversas interrogantes: ¿cómo la tecnología puede apoyar eficientemente estas nuevas estructuras sociales?, ¿cómo escoger la alternativa adecuada? o ¿cómo determinar si una estructura social utiliza la tecnología adecuada?

Encontrar respuestas a estas interrogantes no es simple, no sólo por lo cambiante de la tecnología, sino que además debido a la naturaleza evolutiva de estas estructuras sociales, lo cual provoca que sus necesidades cambien a través del tiempo. Las decisiones que se tomen a lo largo de la vida de la comunidad virtual, u otro tipo de estructura social, serán fundamentales para asegurar su correcto crecimiento.

El fenómeno anterior se vuelve complejo, si consideramos el hecho de que muchas de las comunidades existentes son administradas por sus mismos creadores. Los cuales, por lo general, no siguen una estrategia específica para guiar el crecimiento y evolución de la comunidad, ni tampoco cuentan con las herramientas adecuadas que les permitan tener una visión completa de lo que va sucediendo al interior de la comunidad. Por lo tanto, se ven imposibilitados de poder tomar las acciones necesarias para satisfacer las necesidades cambiantes que se van generando al interior de la comunidad virtual. De esta forma, los administradores pasan a ser meros espectadores de los acontecimientos que suceden en la comunidad, tanto los relacionados con el crecimiento de la misma, como de su posterior extinción.

¿Por qué sucede el escenario anterior? Principalmente por el hecho de que cualquier persona puede comenzar una comunidad virtual, dada la creciente oferta de sitios Web que ofrecen dicho servicio. El creador de la comunidad, en algún momento de su desarrollo se verá enfrentado al problema de tener que tomar decisiones que favorezcan el desarrollo de la comunidad, sin embargo no se encontrará preparado para ello. Dado que la tecnología condiciona las interacciones entre las personas, una mala decisión que se tome limitará las posibilidades de la comunidad virtual que se desea apoyar. Una consecuencia directa de esto es que sus miembros no puedan interactuar de forma adecuada, lo cual provocará la desaparición paulatina de dicha comunidad.

Ahora bien, Wenger [2] señala que en general una comunidad virtual no se crea, sino que se cultiva, al igual como sucede con una planta. Para su crecimiento, es necesario proveerle todos los cuidados necesarios, dándole todos los elementos que requiera, y protegiéndola de todos los peligros existentes. Esto implica una constante preocupación por el crecimiento adecuado de la comunidad, de tal forma que su crecimiento sea lo más saludable posible. En una comunidad virtual este punto toma mayor importancia, puesto que no contar con las herramientas o estrategias adecuadas puede significar tomar decisiones equivocadas al momento de soportar una estructura social.

En general, un administrador, más que conocer las respuestas a las interrogantes mencionadas anteriormente, lo que conoce son las posibilidades que existen en la Web. Por una parte, existen los grandes sitios Web de redes sociales en línea, tales como Facebook o Twitter, los cuales permiten crear grupos en su interior. Por otra parte, existe un gran conjunto de software libre que permite crear comunidades de temas específicos. Con el desarrollo del software libre, ya no necesario desarrollar grandes sistemas para soportar una comunidad virtual, ni es necesario volver a implementar funcionalidades ya existentes. Sin embargo, aún quedan diversas dudas, ¿qué sistema es más apropiado para una cierta comunidad?, ¿cuál es el impacto de utilizar un sistema u otro? En general el problema de tomar estas decisiones recae en el administrador. Si un miembro de una comunidad no se siente satisfecho, dado que su participación es voluntaria, se irá a otra comunidad. El administrador debe lidiar con los nuevos requerimientos que surgen al

interior de la comunidad, teniendo que tomar muchas veces diversas decisiones, las cuales pueden tener consecuencias positivas o negativas al interior de la comunidad.

El desafío planteado en este escenario se debe a que la solución no es meramente tecnológica. El contexto en el cual será utilizada esta tecnología Web, debe ser considerado. Para ello, técnicas o estrategias que no consideren los aspectos sociales de estas estructuras no podrán ser aplicadas satisfactoriamente, dado que este tipo de sistemas son utilizados por estructuras sociales.

1.1 Descripción del Problema a Abordar

La información que se genera al interior de la comunidad es muy amplia y variada, por lo cual su análisis se vuelve complejo y poco factible de realizar por parte de los administradores, quienes generalmente no cuentan con los conocimientos adecuados sobre técnicas de análisis de redes sociales. La mayor parte de las herramientas con que cuenta un administrador de una comunidad virtual no consideran los aspectos sociales de la misma, y las que si lo hacen, analizan los datos generados desde la perspectiva de una red social, es decir, se basan en las relaciones sociales que establecen los miembros de la comunidad al momento de interactuar. Lo anterior ocasiona que se omitan las características propias de una comunidad virtual, como por ejemplo, el hecho de que las relaciones sociales que se forman, a partir de la participación e interacción con otros miembros de la comunidad, se deben a la existencia de un propósito común y que se encuentran reguladas por las políticas de moderación de la comunidad.

A pesar de existir varias teorías reconocidas en el área de sociología, las cuales identifican diversos aspectos sociales que caracterizan a una comunidad virtual [3][2][4][5][6][7], y que existe la necesidad de contar con un proceso cíclico que permita comprender cómo la tecnología es realmente utilizada por una estructura social [8], poco se ha hecho desde el punto de vista computacional para determinar cómo evaluar dichos aspectos sociales, en escenarios donde un sistema de software actúa como mediador de las interacciones. La mayor parte de los trabajos existentes, los cuales se detallan más adelante, se enfocan en evaluar aspectos desarrollados al interior de otro tipo de estructuras sociales, como por ejemplo, las redes sociales en línea. Hay muchos trabajos también que se enfocan en analizar y evaluar las interacciones particulares que se desarrollan al interior de una estructura social, más que en evaluar la estructura social como un todo, el cual tiene características propias.

Lo anterior ocasiona que los administradores, a pesar de invertir mucho tiempo en observar e intervenir en las interacciones entre los miembros de la comunidad, no tengan una visión global de lo que sucede con ésta. Ellos solamente cuentan con la visión parcial que les otorga su participación en la comunidad, y un conjunto de indicadores sobre el uso del sistema de software utilizado, el cual no refleja si la comunidad virtual que se está desarrollando crece adecuadamente o no. Todas esas limitaciones adquieren especial importancia al momento de tomar decisiones que involucren a la comunidad virtual y sus miembros, dado que los administradores no dispondrán de información útil sobre el comportamiento de la comunidad virtual que administran.

1.2 Objetivos de la Tesis

El objetivo general de este trabajo de tesis es proponer una estrategia para realizar el procesamiento semi-automático de aspectos sociales específicos, que desarrolla una comunidad virtual al utilizar un sistema de software como mediador de las interacciones entre sus miembros. Estos aspectos son *personas, propósito y políticas*, los cuales son considerados a través de diversos indicadores. Particularmente se espera aplicar la estrategia propuesta sobre dos comunidades virtuales en las cuales exista por lo menos un administrador que conozca la evolución y funcionamiento de la comunidad virtual, de tal forma que pueda utilizarse su conocimiento para evaluar los resultados obtenidos.

Los objetivos específicos que se desprenden del objetivo general son los siguientes:

1. Especificar la estrategia de procesamiento de la información, en base a algoritmos de minería de datos, análisis de redes sociales y uso de encuestas. Esta estrategia permitirá medir, en forma semi-automática, un conjunto de aspectos sociales que caracterizan a una comunidad virtual.
2. Aplicar la estrategia propuesta en comunidades virtuales reales, de tal forma de evaluar la utilidad y veracidad de los resultados obtenidos.

1.3 Metodología Utilizada

En el presente trabajo se utilizó como base el proceso CRISP-DM (Cross Industry Standard Process for Data Mining) [9], el cual provee un marco de trabajo para analizar la información generada por un sistema de software determinado. Las actividades consideradas en este trabajo de tesis son las siguientes:

1. Revisión del estado del arte relacionado con el análisis de aspectos sociales desarrollados en comunidades virtuales. El énfasis principal de esta revisión está en analizar los aspectos sociales relacionados con el concepto de sociabilidad y las técnicas de evaluación existentes.
2. Recolección de datos históricos desde las comunidades virtuales en estudio, utilizando una estructura común para representar la información de la comunidad.
3. En base a la estructura definida, se determinan las técnicas específicas de evaluación que serán usadas para obtener los indicadores antes mencionados.
4. Aplicación de las técnicas definidas, para obtener los indicadores que apoyan la administración de la comunidad considerando su estado actual.
5. Evaluación de los resultados obtenidos y generación de los reportes correspondientes.

Para el presente trabajo se utilizaron dos comunidades virtuales, tanto para la recolección de información histórica, como para realizar los experimentos que permitieron validar la utilidad y veracidad de los resultados obtenidos. Las comunidades a las que se hace referencia son las siguientes:

- *Comunidad Virtual I:* Se encuentra formada por un grupo de personas reunidas con el objetivo de construir efectos de música, amplificadores y equipos de audio. Ellos se caracterizan por estar orientados hacia “hazlo tú mismo”. En la actualidad, la comunidad cuenta con más de 2.500 miembros, los cuales utilizan un sistema de foro de discusión. En sus más de 10 años de experiencia, sus miembros han compartido y discutido su conocimiento en torno a la construcción de sus propios plexis y efectos. Sin embargo, han aparecido nuevos temas, tales como la lutería, audio profesional, compra/venta de partes, etc. En un comienzo, la administración era una tarea más bien sencilla, efectuada por un solo miembro; sin embargo hoy dicha tarea es realizada por varias personas debido al tamaño de la comunidad. La visión de los administradores y miembros expertos acerca de la comunidad, está basada mayormente en la experiencia y el tiempo durante el cual han participado en la comunidad. Ellos poseen algunas medidas globales básicas que les permiten comprender el estado de la comunidad, como por ejemplo, utilizan el número total de publicaciones, y miembros conectados, entre otros. Sin embargo, ellos no poseen información acerca el comportamiento de los usuarios, calidad de las publicaciones de estas personas, y cómo ellos contribuyen a los objetivos de la comunidad. Para realizar este trabajo se tuvo acceso a la base de datos del sistema de foro de discusión, acceso a los logs del servidor Web, se pudo hacer encuestas a los miembros, y entrevistas a los administradores de la comunidad.
- *Comunidad Virtual II:* Esta comunidad se encuentra al interior de una empresa, y fue creada inicialmente con el objetivo de promover el intercambio de ideas para apoyar la innovación en dicha empresa. La comunidad cuenta con más de 500 miembros en sus más de 3 años de experiencia, los cuales utilizan un sistema Web desarrollado especialmente para la comunidad. En dicho sistema sus miembros pueden subir sus ideas de innovación, y el resto de los miembros puede opinar al respecto. La tarea de administración es realizada por un funcionario contratado especialmente para dicha actividad, el cual sólo utiliza medidas globales básicas, como por ejemplo, el número total de publicaciones, y número de comentarios, entre otros. Se espera que el resto de la organización utilice el sistema (se agregarían 1.500 miembros más), sin embargo, no cuentan con indicadores que les permitan gestionar la comunidad de forma apropiada. Para este trabajo se tuvo acceso a la base de datos del sistema de software utilizado, a los logs del servidor Web de los últimos meses, se contó con la posibilidad de realizar encuestas a los miembros de la comunidad, y entrevistas al administrador.

1.4 Contribución de la Tesis

En el marco de la resolución del problema planteado, este trabajo de tesis espera contribuir con:
(a) una estrategia para evaluar los aspectos sociales definidos anteriormente, específicamente los

relacionados con el concepto de sociabilidad, (b) una guía respecto a cómo interpretar los resultados obtenidos de la evaluación.

Debido a que la mayor parte de los trabajos existentes, los cuales se encuentran detallados en el Capítulo 2.2, se enfocan principalmente en analizar los aspectos que motivan la participación de los miembros de una comunidad, más que el análisis de la estructura social que emerge, se espera además que las soluciones que se propongan en el ámbito de este trabajo ayuden a desarrollar más este enfoque por parte de futuras iniciativas.

1.5 Estructura de la Tesis

La estructura de este documento de tesis considera los siguientes capítulos:

Capítulo 2: Está dedicado a presentar el marco teórico y los trabajos relacionados, incluyendo un resumen de los aspectos sociales involucrados en los principales tipos de estructuras sociales, un resumen de las diferentes estrategias y técnicas de evaluación usadas en comunidades virtuales, y una descripción de los modelos de conceptos sociales existentes.

Capítulo 3: Presenta una visión general de la estrategia de procesamiento desarrollada, mostrándose los aspectos sociales en los cuales se centra.

Capítulo 4: Presenta en detalle las técnicas de evaluación utilizadas sobre los aspectos sociales de *personas, propósito y políticas*.

Capítulo 5: Explica los detalles del diseño y la implementación de la estrategia de procesamiento propuesta.

Capítulo 6: Muestra la aplicación de la estrategia de procesamiento sobre una comunidad virtual existente, específicamente sobre la Comunidad I. Incluye el análisis de los principales resultados obtenidos en conjunto con los administradores de la comunidad en estudio.

Capítulo 7: Muestra la aplicación de la estrategia de procesamiento sobre una comunidad virtual existente, específicamente sobre la Comunidad II. Incluye el análisis de los principales resultados obtenidos en conjunto con los administradores de la comunidad en estudio.

Finalmente se exponen las conclusiones de este trabajo, las referencias bibliográficas consideradas y el material anexo que apoya lo expuesto en los capítulos anteriores.

Capítulo 2: Marco Teórico y Trabajos Relacionados

A continuación se presenta el marco teórico y los trabajos relacionados con esta tesis. La descripción de dichos trabajos permitirá un mejor entendimiento de la propuesta presentada en este documento de tesis.

2.1. Estructuras Sociales

La masificación del uso de servicios basados en Internet, tales como los sistemas Web, ha permitido que muchas personas puedan comunicarse e interactuar con otras sin importar su ubicación geográfica. Es posible conversar con otras personas, buscar a gente con intereses comunes, ayudar a otros en determinados problemas, compartir información, y participar en debates, etc. Estas actividades han provocado que el uso del computador pase de ser una actividad individual, a una actividad colectiva en la cual se crean diferentes nexos de interacción y cooperación con otras personas.

Dado este escenario, el uso de Internet se vuelve cada vez más importante, puesto que ha permitido que emerjan nuevas estructuras sociales [10] [11] las cuales, a pesar de estar basadas en estructuras sociales ya existentes, poseen características propias [3] que deben ser consideradas al momento de efectuar un estudio que las involucre. Estas diferencias son debido al medio de interacción que utilizan, el cual ya no es cara-a-cara, lo cual provoca que muchos de los rituales sociales que existen en el mundo real [12] no existan o estén limitados en el mundo virtual [13].

De esta forma, emergen nuevas formas de relaciones entre las personas, a través de estas nuevas estructuras sociales: redes sociales, comunidades virtuales, comunidades virtuales de práctica, etc. Comprender cuáles son los principales aspectos de estas estructuras sociales es fundamental para evaluarlas adecuadamente. En este capítulo se presentan los aspectos sociales más importantes de estas estructuras. Esto permite comprender, tanto el contexto social en el cual se desarrollan estas estructuras sociales, como los aspectos que los sistemas de software deben incluir al momento de evaluar su funcionamiento.

2.1.1. Redes Sociales

El uso de Internet ha posibilitado conectar no solamente computadores a través de la red, sino que también ha permitido conectar a las personas [14] [7]. El uso del correo electrónico, foros de discusión, y otros sistemas han permitido que las personas ya no tengan que trabajar en forma aislada, sino que han podido comenzar a trabajar en grupos en línea, facilitados por la formación de redes sociales.

Una red es básicamente un conjunto de objetos (llamados nodos) y un conjunto de relaciones entre estos objetos. Basado en lo anterior, una red social es un conjunto de actores sociales relacionados a través de vínculos sociales, como por ejemplo amistad o trabajo conjunto. Estos actores pueden ser desde personas individuales, grupos u otro tipo de organización social. Una red puede ser representada gráficamente a través de un sociograma (Figura 1). En dicha representación, se utilizan puntos para indicar a las personas y líneas para representar las relaciones, las cuales pueden ser sólidas o intermitentes, de acuerdo a si las relación es fuerte o débil.

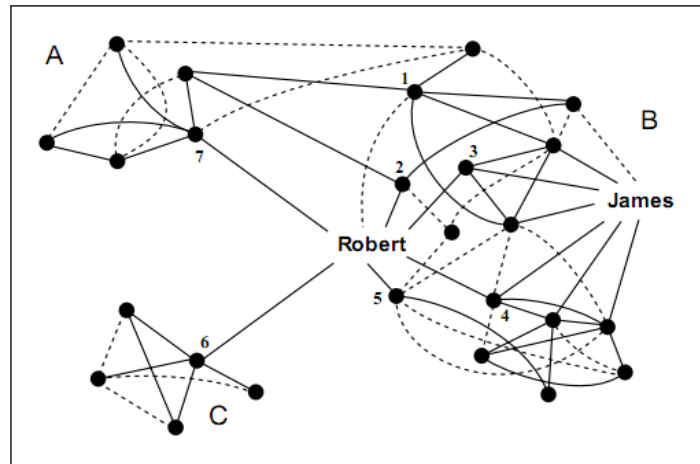


Figura 1. Representación de un sociograma.

Las redes sociales se han popularizado en los últimos años por la existencia de plataformas Web como Facebook o Twitter, entre otros. En ellas, una persona puede comunicarse con otras, creando relaciones de tipo laboral, familiar, estudio, o amistad, entre otras. De esta forma, las relaciones entre los distintos usuarios de estas plataformas van conformando una gran red social. Bajo la perspectiva de una red social, los siguientes aspectos son fundamentales:

1. Los actores y sus acciones son vistos como interdependientes, en vez de verlos como unidades independientes y autónomas.
2. Las relaciones existentes entre los actores son canales que permiten la transferencia o flujo de recursos (ya sean éstos tangibles o no) [15].

Un grupo de personas que interactúa en línea construye implícitamente una red social, aunque ésta no sea definida explícitamente [14]. Por lo tanto, dado que una persona puede participar en varios grupos, una misma persona puede ser parte de varias redes sociales.

Las diferencias entre una red social en línea con una red social real (que no es en línea), surgen básicamente por el mecanismo a través del cual interactúan dos actores dentro de la red: en uno será a través de un medio virtual y en otro será a través del mundo real. Gross [16] señala 3

grandes diferencias entre estas redes: (i) mientras que en el mundo real, una relación social puede ser percibida por una persona de muy diversas formas, desde ser muy débil a muy fuerte, en una red social en línea tienden a ser reducidas a simples relaciones binarias de “es mi amigo o no lo es”. Esta situación lleva al hecho de que una persona indique como su amigo a alguien a quien apenas conoce realmente [17]; (ii) mientras que la cantidad de nexos fuertes que una persona puede mantener en una red social en línea no aumenta significativamente con el uso de tecnología, el número de nexos débiles si puede incrementarse sustancialmente, dado que el tipo de comunicación que existe es más adecuada para ese tipo de nexos [18]. Finalmente, (iii) mientras que en una red social real una persona podría tener docenas de nexos significativos y entre 1.000 y 1.700 conocidos, en una red social en línea es posible encontrar cientos de “amigos” y cientos de miles de conocidos que se encuentran a no más de 3 grados de separación de la persona.

Lo anterior implica que una red social en línea sea mucho más amplia y tenga muchas más conexiones débiles que una red social real [16]. Dicha situación provoca un desafío diferente, relacionado con la privacidad de las conexiones que posee una persona. Dado que una persona es mucho más “conocida”, su información será también conocida por un mayor número de personas. Considerando lo anterior, ¿qué hace que una persona quiera hacer pública sus conexiones con otras personas? Para Burt [19] el camino más básico para establecer confianza en nuevas relaciones con personas, es que éstas ya sean conocidas por gente en las que uno ya confía. De esta forma, las conexiones que posee una persona constituyen una forma de desplegar y aumentar su capital social [18], el cual es reconocido por otras personas de la red social.

El capital social se refiere a la idea general de que participar en redes sociales u otras estructuras sociales puede tener consecuencias positivas para un individuo [20]. Dicho capital es intangible y es un mecanismo a través del cual las personas adquieren conocimiento [21]. Putnam [22][23], basado en la evidencia, señala que existen tres formas de capital social [24]: (a) *Capital de Red*: compuesto por las relaciones que posee una persona, con sus amigos, vecinos, parientes, compañeros de trabajos, etc., y todas aquellas relaciones que proveen algún beneficio, como compañía, ayuda emocional, bienes o servicios y sentido de pertenencia [25]; (b) *Capital Participativo*: la participación en organizaciones, de forma voluntaria, ofrece la oportunidad a las persona de crear logros en conjunto, articulando sus demandas y deseos, y (c) *Compromiso Comunitario*: el capital social consiste en mucho más que interacciones entre personas y participación en una organización social. Cuando las personas tienen una actitud fuerte hacia una comunidad, ellos movilizarán su capital social de buena gana y de forma efectiva. Por lo tanto, conociendo las conexiones de alguien, no sólo es posible determinar información de esa persona relacionada con sus amistades, gustos musicales u otra información similar, sino que también información relacionada a cuán importante es la persona dentro de la red social. Ahora bien, ser miembro de una red social tiene otras ventajas, tales como: (a) permite conocer las necesidades actuales de los demás y estimula estas relaciones a través de contactos más frecuentes en el tiempo [26]; (b) permite intercambiar fotos, canciones y otros tipos de archivos; y (c) permite conocer personas.

Para caracterizar una red social en línea es posible hacerlo a través del análisis de las diversas propiedades que posee, por ejemplo, a través de sus niveles de reciprocidad, densidad y formación de componentes a través del tiempo [27]. Estudios previos [27] [28] [29] han mostrado

que grandes redes sociales en línea, como Flickr, Youtube, LiveJournal, Yahoo! 360 y Orkut, a pesar de ser diferentes, poseen valores similares de algunas propiedades mencionadas anteriormente, tales como la reciprocidad (entre un 62 y 100%). Este hecho muestra un hecho interesante, relacionado con el crecimiento similar de las redes sociales en línea, y la similitud en la forma en que las personas interactúan en las redes sociales.

2.1.2 Comunidades Virtuales

Buscar una definición precisa de comunidad no es una tarea sencilla. Es posible encontrar numerosas definiciones, desde una perspectiva social, las cuales han ido cambiando y se han redefinido en el tiempo [30] [31]. Sin embargo, existen indicadores de comunidad que han sido adoptados por diversos investigadores: el concepto de personas compartiendo algún interés común, experiencias y/o necesidades, enlazados a través de relaciones sociales, a través de los cuales se obtienen recursos importantes, desarrollan fuertes sentimientos interpersonales de pertenencia y necesidad mutua, y se da el surgimiento de un sentido de identidad compartida [32][33][34].

Los tipos de comunidades que se crean a través de la red difieren en varios aspectos respecto al concepto tradicional de comunidad, principalmente por el hecho de que son mediadas a través de computadores. A estos tipos de comunidades se les denomina “comunidades virtuales” o “comunidades online”. En una comunidad virtual las personas pueden o no encontrarse cara a cara con otras, y el intercambio de mensajes e ideas se efectúa a través de la red [33]. Una comunidad virtual existe y juega un rol socializador de la misma forma que lo hacen las comunidades “reales” [8] [33].

Las comunidades virtuales pueden ser vistas como una red social, dado que la red de computadores que conecta a las personas permite crear un conjunto de enlaces con un significado social [7] entre las personas. De la misma forma, un grupo de trabajo también es considerado una red social, siendo la principal diferencia con una comunidad virtual en que las relaciones entre las personas se encuentran delimitadas firmemente en su frontera (no existen relaciones con personas que no pertenecen al grupo de trabajo) y son muy densos (cada persona se encuentra relacionada con la mayoría de los miembros del grupo) [7]. Una comunidad virtual se crea a partir de la continuidad de las relaciones entre sus miembros, pero es experimentada a través de actividades específicas localizadas en el tiempo y el espacio [8]. Los principales tipos de comunidades que se estudian en la actualidad son los siguientes:

1. *Comunidades de interés* [35]. Son aquellas comunidades en las cuales sus integrantes comparten el mismo interés en algún tópico (y por tanto todos ellos poseen un background común). Ejemplos de este tipo de comunidad son: los fan club de grupos musicales, los grupos de personas interesadas en los planetas del sistema solar, etc. Otro tipo de comunidades son las llamadas comunidades de pasión [36], las cuales poseen una definición muy similar.

2. *Comunidades de práctica* [37][8][38]. Son aquellas comunidades en las cuales sus integrantes comparten una misma profesión o actividad. Generalmente sus miembros se encuentran fuertemente involucrados en la comunidad. Ejemplos de este tipo de comunidad son: la comunidad de programadores de *Java*, la comunidad *Open Source*, comunidades dentro de empresas, etc.
3. *Comunidades de propósito* [36]. Son aquellas comunidades en las cuales sus miembros comparten el mismo objetivo (de corto plazo generalmente). Ejemplos de este tipo de comunidad son los compradores de una librería virtual, los cuales comparten el objetivo de encontrar y comprar un libro. Los miembros tienen un propósito funcional que es disuelto una vez que el objetivo es alcanzado. Generalmente los miembros de este tipo de comunidad no realizan actividades que excedan los propósitos de la comunidad ni comparten necesariamente el mismo interés [36].

Existen otros tipos de comunidades, las cuales no se detallan en este trabajo. Una revisión de éstas puede ser encontrada en [30][4].

Henry & Pudelko [4], basados en la teoría de Wenger sobre conocimiento [39], establecen una relación entre los distintos tipos de comunidades que podemos encontrar. Ellos establecen que existe una relación entre el tipo de comunidad que se forma y la cohesión e intencionalidad de la comunidad misma (Figura 2). Por ejemplo, en una comunidad de interés se presenta la menor cohesión posible y la intencionalidad, al momento de la creación de la comunidad, es leve. En el otro extremo, en cambio, encontramos a las comunidades de práctica, en las cuales el grupo de personas que la componen se encuentran fuertemente cohesionadas, y en donde la formación de la comunidad es altamente intencionada. A continuación describimos las comunidades de prácticas que son el objeto de estudio de esta tesis.



Figura 2. Tipos de comunidades virtuales (Figura obtenida de [4]).

2.1.3 Comunidades Virtuales de Práctica

Las Comunidades Virtuales de Práctica (VCoP por su sigla en inglés) han experimentado un crecimiento explosivo en Internet en los últimos años. El valor que posee este tipo de comunidades es que permiten establecer relaciones entre personas que quieren compartir o aprender acerca de un tema específico, basados en la interacción de dichos miembros [39]. El rol que posee Internet es fundamental, puesto que facilita la interacción entre los miembros de la comunidad sin necesidad de contacto presencial, como sería el caso de una comunidad “real”. Para poder soportar este tipo de interacciones la Web ha sido fundamental, ofreciendo un medio de interacción mucho más significativo en comparación con otras tecnologías como el correo o la mensajería instantánea. Las herramientas más comunes utilizadas a través de la Web son los foros, wikis, y otras herramientas similares.

Para una VCoP es muy importante poder generar, almacenar y mantener el conocimiento resultante de la interacción entre sus miembros. El éxito de una VCoP depende en gran parte de los mecanismos de administración [40] y de las participaciones de los “miembros claves” de la comunidad (también llamados líderes [41] o miembros principales [40]). Del mismo modo, el objetivo de cada miembro de la comunidad virtual de práctica es aprender cierto conocimiento de la comunidad, por lo tanto, el contenido de las interacciones es un aspecto fundamental que debe ser considerado en este tipo de estructuras sociales.

Plaskoff [42] ha sugerido que para que exista una comunidad de este tipo, basado en su propia experiencia en una gran compañía farmacéutica, deben existir 3 conceptos, de los cuales depende el éxito de poder cultivar una comunidad de práctica en una organización: *believing*, *behaving*, y *belonging*. *Believing* se refiere a la idea de que los miembros deben creer en el valor intrínseco de la comunidad. *Behaving* indica que los miembros forman y siguen normas al interior de una comunidad. *Belonging* significa que cada miembro cultiva un sentido de pertenencia con la comunidad.

Por otro lado, para que una comunidad de práctica pueda funcionar a través de Internet, primero se necesita crear una comunidad virtual. Sin embargo, la creación de una comunidad virtual no garantiza, necesariamente, que una comunidad de práctica va a ser desarrollada, puesto que se necesita que exista un esquema de aprendizaje basado en tareas [13]. Existen 3 componentes que hacen que una estructura social sea considerada como una “comunidad de práctica” (CoP) y no sólo como una “comunidad”: *dominio*, *comunidad* y *práctica* [37]:

1. *Dominio*. Una comunidad de práctica no es sólo un club de amigos o una red de conexiones entre personas. Tiene una identidad definida por un dominio compartido de intereses. Ser miembro de la comunidad implica un compromiso, y por tanto una competencia (responsabilidad) compartida que distingue a los miembros de otras personas [37].
2. *Comunidad*. Al perseguir sus intereses en su dominio, los miembros se interesan en participar en actividades y discusiones, ayudar a otros y compartir información. Ellos

construyen relaciones que les permiten aprender de los demás. Es necesario "interactuar" y "aprender de otros" para tener una comunidad de práctica [37].

3. *Práctica*. Una CoP no es sólo una comunidad de personas con intereses quienes tienen un gusto en común. Los miembros de una CoP son prácticos, ellos desarrollan un repertorio compartido de recursos: experiencias, historias, herramientas, formas de direccionar problemas recurrentes, etc. El desarrollo de prácticas compartidas puede ser más o menos un proceso consciente.

Un aspecto fundamental de las comunidades de práctica es el hecho de que éstas evolucionan con el tiempo [39]. La vida de una comunidad es un viaje de descubrimiento que le ayuda a reinventarse continuamente [8] [6] [43]. En la evolución de una comunidad es posible identificar 5 etapas (*ver Figura 3*): *Descubrimiento, Incubación, Expansión, Preservación y Transformación* [2].

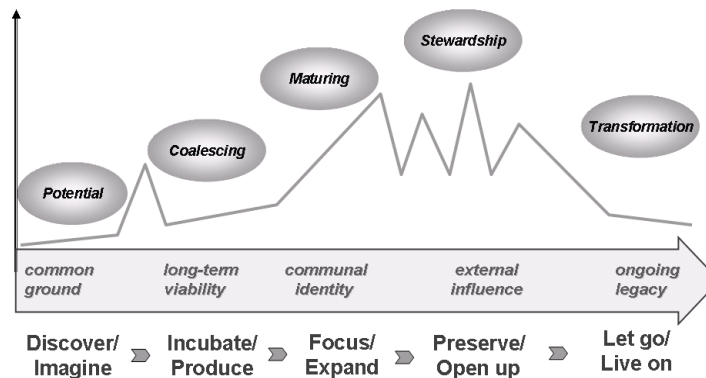


Figura 3. Evolución de una comunidad de práctica (Fig. obtenida de [2]).

De la misma forma, la participación de los miembros de una comunidad puede evolucionar con el tiempo, desde ser observadores, hasta ser participantes ocasionales o líderes y coordinadores de la comunidad. Wenger define una serie de niveles de participación de los miembros de una comunidad [37], a partir de las relaciones que forjan con los demás miembros (*ver Figura 4*).

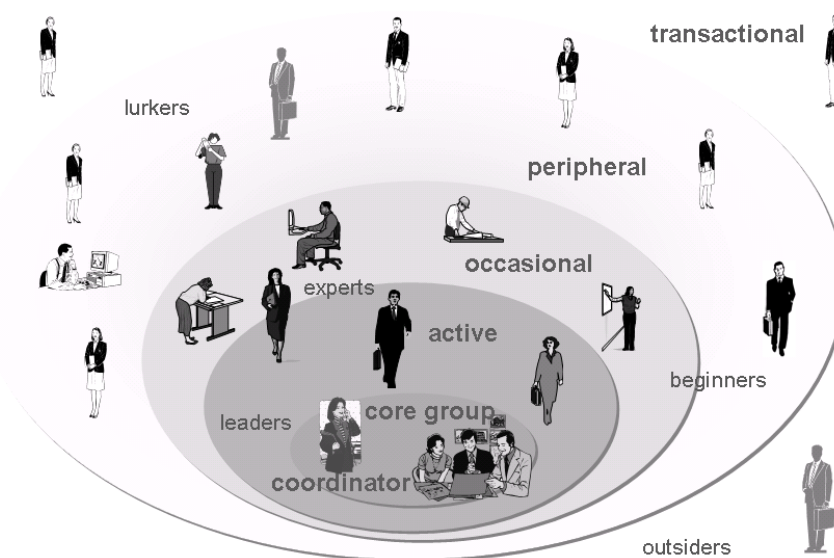


Figura 4. Niveles de participación de los miembros de una comunidad (Figura obtenida de [2]).

Como se señaló anteriormente, una VCoP comparte características con otras estructuras sociales, específicamente con las comunidades virtuales y redes sociales en línea, dado que éstas corresponden a estructuras mucho más genéricas en las cuales se basa la definición de una VCoP. Este hecho provoca que el análisis de este tipo de estructuras sociales sea más complejo respecto de otras más genéricas (redes sociales por ejemplo), dado que será necesario estudiar tanto los aspectos que comparte con estas estructuras sociales, así como los aspectos propios que posee.

A continuación, en la Tabla 1, se muestra un resumen de los aspectos sociales más importantes involucrados en cada una de las estructuras sociales mencionadas anteriormente, las cuales sirven como una guía para el análisis de una comunidad virtual de práctica. El detalle de los aspectos sociales mencionados es posible revisarlos en el Anexo B.

Tabla 1. Resumen de los aspectos sociales de las estructuras sociales en estudio, desde la perspectiva de diversos autores.

Tipo de Aspecto Social	Aspectos Sociales
Personas y relaciones sociales que se establecen.	Personas [44] [45], Actores [46], Roles [3] [7], Liderazgo [3], Participación [47], Grupos [7], Subgrupos [3], Identidad [47], Belonging [42], Relaciones [46] [48], Lazos [46], Perfiles [3].
Propósito de participar en la comunidad.	Propósito [44], Intención [4], Propósito [3], Dominio [39], Contenido Común [48].
Políticas y formas de interactuar.	Políticas [44], Etiquete [3], Integración [47], Behaving [42], Autonomía [47].

2.2 Estrategias de Evaluación de Comunidades Virtuales

Es posible analizar el funcionamiento y comportamiento de una comunidad virtual desde diversos puntos de vista. Los enfoques más clásicos se orientan al análisis del contenido generado por los miembros de la comunidad al interactuar a través del tiempo, sin considerar que existe una estructura social que interactúa a través de un sistema de software. Por ejemplo, a través del análisis de la calidad de las respuestas que se dan al interior de una comunidad [49], el análisis del comportamiento de los comentarios generados por los miembros de una comunidad [50], o el análisis del sentimiento u opinión de un miembro de la comunidad acerca de un tema en particular [51]. El problema de este enfoque es que, al evaluar únicamente el resultado de la interacción entre los miembros de una comunidad virtual, se dejan de lado los aspectos que se desarrollan al interactuar los miembros de la estructura social. Por lo tanto, se ignora la información de la estructura social que emerge como consecuencia de dichas interacciones.

Contrario al caso anterior, se encuentran los enfoques más modernos que sí consideran los aspectos de la estructura social que emerge como resultado de las interacciones entre las personas. Es posible distinguir dos enfoques: (a) *el psicológico*: el cual tiene un énfasis en el estudio de los factores humanos que influyen en la participación de los usuarios en una estructura social en línea, y (b) *el sociológico*: el cual tiene un énfasis en analizar la estructura social que se forma producto de las interacciones entre los miembros de una comunidad, considerando a dicha estructura social como una entidad con sus propias características y formas de funcionamiento. A continuación se detallan estos dos enfoques, indicando ejemplos de trabajos que utilizan dichos enfoques.

2.2.1 El Enfoque Psicológico

Este enfoque se basa en teorías tales como la teoría cognitiva social, la teoría de capital social, la teoría de redes sociales y la teoría del compromiso [52]. Entre los trabajos relacionados que utilizan enfoques psicológicos se encuentran, por ejemplo, trabajos que se orientan en comprender los factores que motivan a una persona a continuar participando en una red social [53] o a compartir conocimientos específicos con otros miembros [54], o la reacción que ocasiona la participación de un miembro sobre otros, basados en la estimación del beneficio percibido por los miembros de una estructura social virtual al participar y la influencia que puede ocasionar la diferencia de género en el beneficio percibido [55]. También hay trabajos basados en la teoría de balance estructural de Heider [56], al medir el beneficio percibido por otros usuario o grupos de usuarios [57].

Existen otros trabajos que se enfocan en el estudio de roles específicos que emergen al interior de la estructura social, como por ejemplo los observadores de una comunidad virtual [58] o los líderes y los tipos de liderazgo que emergen al interior de una comunidad [59], y la forma en que éstos influyen en la motivación y participación de otros miembros. Una de las ventajas de estos enfoques es que permiten realizar predicciones acerca del comportamiento futuro de los miembros de una comunidad. Por ejemplo, West et al. [60] predicen la opinión que tendrá un

miembro A acerca de otro miembro B, basado en la mezcla de dos tipos de informaciones: la red social a la que pertenece la persona A y lo que escribe, analizando lingüísticamente los sentimientos que expresa. Por otra parte, Cheng et al. [61] analizan cómo la evaluación entre miembros (a través del uso de rating de los comentarios publicados) afecta el comportamiento futuro de un autor dentro de una comunidad, regulando la calidad y cantidad de contribuciones futuras de sus miembros.

La principal limitación de los trabajos anteriores es que no permiten evaluar el funcionamiento de una comunidad virtual como un todo, dado que su orientación es principalmente hacia el análisis del comportamiento de los individuos que participan al interior de una comunidad virtual o red social en línea.

2.2.2 El Enfoque Sociológico

La mayor parte de los trabajos que utilizan esta estrategia están centrados en el uso de algoritmos de Análisis de Redes Sociales (SNA – Social Network Analysis), los cuales permiten obtener información acerca de la estructura de la comunidad a través del análisis automático de los datos que genera la estructura social. Por ejemplo, es posible encontrar trabajos que analizan la estructura de los usuarios utilizando las propiedades de inequidad de la participación [62], trabajos que buscan definir estrategias para detectar comunidades dentro de una red social [63][64], trabajos que realizan un análisis de las funciones relacionadas a la moderación [65][66], o detectan miembros que cumplen roles específicos, como el de intermediarios [67]. También, dentro de este enfoque, existen trabajos que analizan la dinámica de la estructura social y cómo se propaga la información en la red social. Por ejemplo, en [68] se analizan dos dinámicas: (1) creación y destrucción de la red subyacente de seguidores, y (2) la dinámica del flujo de información a través de la re-publicación de contenidos generados por otros miembros.

Lo que caracteriza a la mayoría de los trabajos antes descritos, es la utilización de técnicas que analizan los logs con la actividad de los miembros de estas estructuras sociales. Si bien hay un mayor énfasis en el uso de técnicas de análisis de redes sociales, es posible ver algunos trabajos que incorporan técnicas de text mining para sus estudios [69]. Trabajos más frecuentes se enfocan en estudiar tipos específicos de comunidades virtuales [70]. Sin embargo, éstos analizan solamente las conexiones entre los miembros de una comunidad, más que las propiedades que la diferencian de otras estructuras sociales [71].

A pesar de que los trabajos anteriores utilizan información acerca de la estructura social que se genera, no obtienen indicadores globales que puedan ser utilizados como apoyo en la toma de decisiones por parte de los administradores. Tampoco consideran aspectos propios de las comunidades virtuales, sino que las analizan como si fueran redes sociales en línea, con lo cual dejan de lado todos los aspectos que caracterizan a una comunidad virtual. Hay también otros trabajos intentan generar dichos indicadores, sin embargo las métricas que se obtienen no dan cuenta de los aspectos sociales que se desarrollan, sino que se enfocan en aspectos particulares del área de estudio. Por ejemplo, en marketing se usan este tipo de indicadores para medir la importancia de una marca [72], o para mejorar las sugerencias de un sistema de recomendación [73] [74]. También se usan para analizar la evolución de lo que siente un grupo de personas

acerca de diversos tópicos de interés [75], para comparar distintos foros utilizados por comunidades virtuales, y determinar así los tipos de miembros que participan [76].

2.3 Técnicas de Evaluación en Redes Sociales y Comunidades Virtuales

El proceso de evaluación es el único que puede garantizar, de una forma científica, el efecto o consecuencia de determinadas acciones que son tomadas en un momento dado. Existen diversas estrategias, desde aquellas que evalúan al software en sí mismo, hasta aquellas que intentan evaluar la estructura social resultante. A continuación se detallan las principales técnicas de evaluación, de acuerdo a las principales áreas involucradas.

2.3.1 Análisis de Redes Sociales

El área de Análisis de Redes Sociales (SNA) tiene como objetivo estudiar estructuras sociales modeladas como redes, es decir, como un conjunto de actores y relaciones sociales entre estos actores. A diferencia de otros tipos de análisis, SNA no es lineal [77], es decir, su análisis no puede ser realizado incrementalmente, sino que debe ser realizado sobre la estructura completa [78]. Esta propiedad de SNA se debe a que está basada en un análisis estructural [79], es decir, el análisis está basado en las relaciones entre los actores y los patrones que surgen en la red producto de estas relaciones. Para realizar un análisis SNA es posible identificar 6 diferentes técnicas [80], entre las cuales se destacan:

1. *Métricas*: El objetivo de utilizar métricas es medir propiedades propias de la red social, de sus actores o un conjunto de ellos. Por ejemplo, es posible medir la densidad de una red social para determinar cuan relacionados se encuentran sus miembros, o es posible medir la centralidad de un actor, para determinar su importancia dentro de la red social. Entre las métricas, podemos encontrar [77] [81]:
 - a. *Centralidad*: El grado en el cual un actor se encuentra en un rol central dentro de la red.
 - b. *Prestigio*: También conocido como estatus, esta medida intenta cuantificar el ranking que posee un actor en particular respecto de un conjunto de actores.
 - c. *Homofilia*: El grado en el cual actores similares en roles similares comparten información.
 - d. *Aislamiento*: Un actor que no posee lazos con otros actores.
 - e. *Puerto*: Un actor que conecta la red con las influencias externas (exterior).

- f. *Punto de Corte*: Un actor cuya remoción resulta en caminos inconexos dentro de la red.

Entre las métricas que son propias a una red, podemos encontrar [77]:

- a. *Centralización*: La fracción de actores principales dentro de la red.
 - b. *Accesibilidad*: El número de vínculos que conectan a los actores.
 - c. *Conectividad*: La capacidad de los actores para llegar de uno a otro recíprocamente, es decir, la capacidad de elegir una relación entre ambas partes.
 - d. *Balance*: La medida en que los lazos en la red son directos y recíprocos.
2. *Subgrupos Cohesivos*: Se refiere a subconjuntos de actores cuyas relaciones son relativamente fuerte, directa, intensa, frecuente o positivas. Entre los métodos para detectarlos, podemos encontrar:
- a. *Completa Mutualidad*: Este método se basa en la búsqueda de los subgrupos cohesivos en los cuales existen todas las relaciones sociales posibles entre todos los actores. Para ellos, se basa en la definición de un *clique*, es decir, un subgrafo maximal completo de al menos tres nodos.
 - b. *Accesibilidad y Diámetro*: Este método se basa en una extensión del método anterior. Estos grupos son importantes si sabemos que los procesos sociales más importantes se producen a través de intermediarios. Se basa en el hecho de que no todos los miembros de un subgrupo se encuentran conectados directamente, sin embargo, las rutas que los conectan son relativamente cortas. Ellos se basan en la definición de n-clique (el cual es el subgrafo maximal de distancia geodésica más larga entre dos nodos cualquiera no mayor a n; una ruta geodésica incluye cualquier nodo dentro del grafo), n-clans (es un n-clique en el cual la distancia geodésica entre todos los nodos en el subgrafo no es mayor que n para los caminos del grafo) y n-clubs (el cual es un subgrafo maximal de diámetro n, es decir, la distancia entre todos los nodos del subgrafo es menor o igual a n)
 - c. *Nodal Degree*: este enfoque está basado en las restricciones sobre el número mínimo de actores adyacentes a cada actor en un subgrupo. Está basado en el concepto de k-plexes (es un subgrafo maximal en el cual cada nodo en el subgrafo le puede faltar no más de k lazos a otros miembros del subgrafo) y k-cores (es un subgrafo en el cual cada nodo es adyacente a por lo menos k nodos dentro del mismo subgrafo)
 - d. *Matrices de Permutación*: El propósito de las matrices de permutación es, dada una red social, representarla como una matriz NxN, reordenando sus columnas y filas con el objetivo de identificar bloques al interior de la red social. Cada fila y columna n corresponde a un miembro, y cada celda indica la existencia o no de un relación social entre dos miembros de la red social [82].

3. *Visualización*: Una red social puede ser modelada como un grafo, representando los actores sociales por círculos y las relaciones por flechas que conectan esos círculos. Esta representación visual es denominada sociograma. Permiten explorar y entender una red social a través del análisis visual de sus relaciones. La complejidad de un sociograma se encuentra en la técnica que se utiliza para dibujar la red social [83], entre las cuales es posible encontrar:
 - a. *Disposición Circular*: Todos los nodos son puestos en un círculo [84].
 - b. *Disposición en Radio*: Los nodos son puestos en círculos concéntricos, en el cual los nodos más centrales son ubicados en el centro del diagrama.
 - c. *Disposición Agrupada*: Es utilizado para desplegar información acerca de grupos, mostrando los nodos cerca de aquellos que comparten alguna característica.
 - d. *Disposición Libre*: en el cual no hay ningún criterio para el despliegue, solamente que sea legible.

2.3.2 Análisis de Comunidades Virtuales

Basados en [85] podemos identificar 4 formas de analizar una comunidad virtual:

1. *Etnografía y técnicas asociadas*: el propósito de la investigación etnográfica es comprender un grupo o situación desde el punto de vista de sus miembros/participantes. Corresponde a un método de investigación cualitativo para comprender cómo la tecnología es usada in-situ [86] [85]. Un ejemplo de esta técnica es la aplicación de un análisis etnográfico a una comunidad virtual que usa una herramienta IRC como medio de comunicación entre sus miembros [87]. Dado que la etnografía, en su forma original, no contempla las dificultades propias de la participación asíncrona y el hecho de que los miembros son virtuales, es que diversos autores han propuesto variantes a la técnica original, tales como la netnografía [88] y la cyberetnografía [89].
2. *Cuestionarios*: son útiles para recolectar información demográfica y tienen la ventaja que ellos pueden ser distribuidos por mano a los participantes locales, o enviados por correo electrónico, o estar en la Web [6] [85]. Sin embargo, a pesar de que los cuestionarios proveen información útil de los miembros de una comunidad [90], ellos no son suficientes. Es recomendable el uso de una técnica secundaria, cuando sea posible, para reducir la subjetividad de la opinión de los miembros de la comunidad.
3. *Experimentos y cuasi-experimentos*: estos estudios son valiosos para evaluar la usabilidad de las interfaces y la reacción de los usuarios a las nuevas características de las interfaces de usuario. Estos enfoques pueden ser perfectamente utilizados para investigar el impacto de cambios en el diseño del software que utiliza una comunidad en línea [6]. Un ejemplo puede ser encontrado en [91]. Para aplicar este enfoque es necesario que los miembros de una comunidad virtual interactúen en un ambiente controlado.

4. *Minería de Datos y Análisis de Redes Sociales*: Este tipo de análisis consiste en utilizar los registros (logs) generados por el software para descubrir información útil de la comunidad virtual. De esta forma, el punto principal es el estudio de la naturaleza social de la comunidad, es decir, el estudio de sus miembros y las relaciones que ellos establecen [6]. En estos estudios típicamente dos estrategias son aplicadas. Primero, aquellas que se encuentran basadas en la visualización de la red social subyacente. Como resultado, es posible deducir o establecer la existencia de patrones de comportamiento u otras estructuras sociales al interior de la red social [92]. Segundo, aquellos que intentan cuantificar de cualquier manera posible las relaciones que son establecidas al interior de la comunidad [93]. Para obtener lo anterior, es común definir y medir un conjunto de métricas e indicadores que permiten monitorear el comportamiento de la comunidad.

2.4 Modelo de Conceptos Sociales

El mundo se puede describir como un conjunto de conceptos y relaciones. Es útil no sólo para tener un entendimiento de los conceptos más importantes, sino que permite facilitar la construcción de sistemas de software. Un modelo de conceptos adecuado permite efectuar desde la creación de los repositorios de datos hasta la creación de los modelos de clases de un sistema de software.

Con el nacimiento de la Web Semántica, el uso de ontologías y meta-datos, contar con una descripción adecuada de los aspectos involucrados en un sistema es fundamental. En la actualidad existen dos modelos que se utilizan frecuentemente: FOAF (Friend of a friend) [94] y SIOC (Semantically-Interlinked Online Communities) [95]. Sin embargo, ambos modelos carecen de algunos aspectos sociales que caracterizan a una comunidad virtual, por lo cual sólo pueden ser utilizados como base, más que como una referencia literal. A continuación se muestran ambos modelos y un modelo propuesto más reciente, denominado SIOC-SNA-DM [96] [97], el cual es una extensión de SIOC y que aborda las características faltantes de la versión original.

2.4.1 Friend of a Friend

FOAF es un vocabulario, expresado en RDF, que permite describir personas y relaciones entre ellas (ver Figura 5). Para ello, una persona “conoce” a otras, de forma recíproca, es decir, si A conoce a B, entonces B conoce a A. Esta relación no especifica su tipo, pudiendo utilizarse para cualquier tipo de relación (amistad, compañeros de trabajo, etc.). FOAF posee la ventaja que se encuentra ampliamente extendido en la Web, por lo tanto ha sido la base de diversas propuestas. La desventaja que posee es que se encuentra orientada más al uso de Redes Sociales que de Comunidades Virtuales, al centrarse en las relaciones existentes entre los miembros de una Estructura Social. En otras palabras, si utilizamos sólo FOAF, estamos enfocándonos en el aspecto social “Personas”, y dejando de lado la existencia de un “Propósito” y de las “Políticas”.

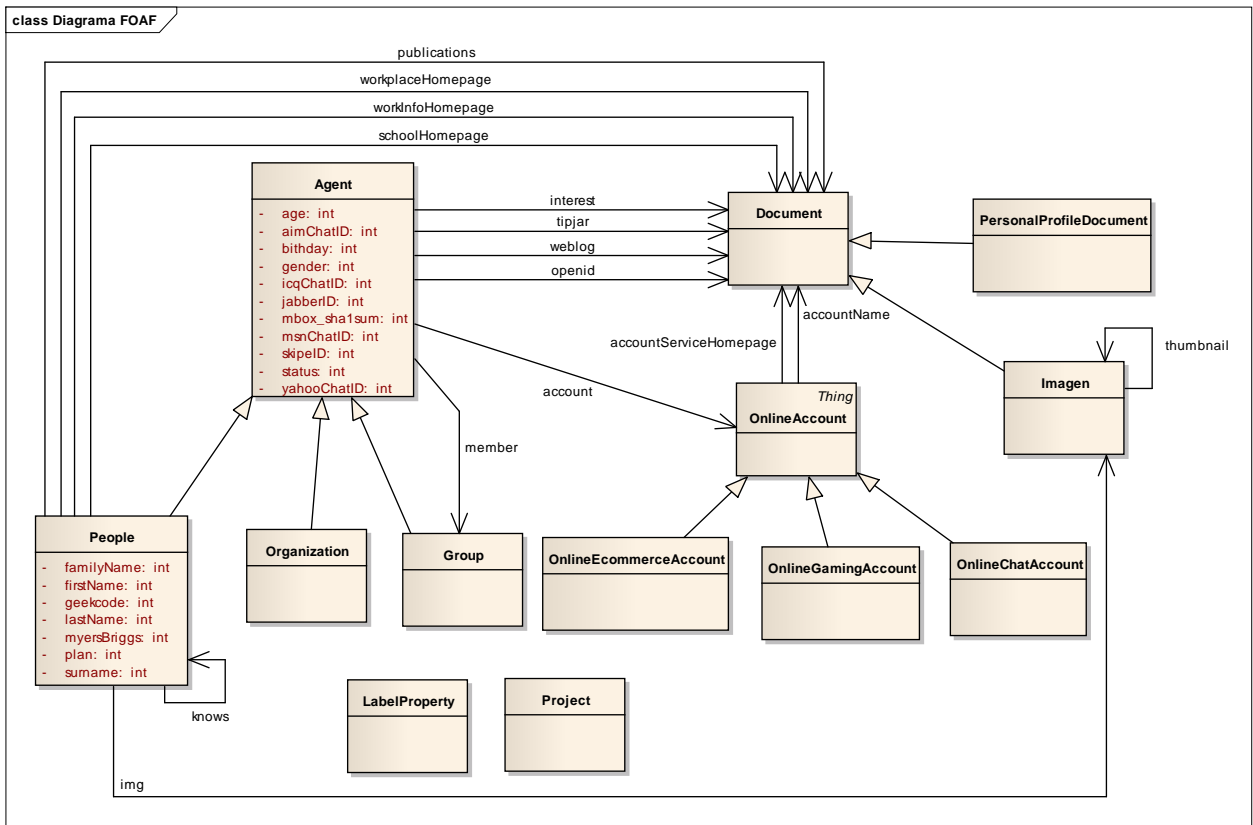


Figura 5. Modelo FOAF.

2.4.2 Semantically-Interlinked Online Communities

SIOC (Semantically-Interlinked Online Communities) [95] corresponde a un vocabulario, también expresado en RDF, que permite describir formalmente los principales conceptos y propiedades que describen la información de una comunidad virtual: foros de discusión, listas de correo, wikis, blogs, etc. El objetivo de que se encuentre en RDF es proveer información que sea legible en forma automática por otros sistemas. De esta forma, toda la información que se genera al interior de una comunidad virtual puede ser expresada con este vocabulario (ver Figura 6).

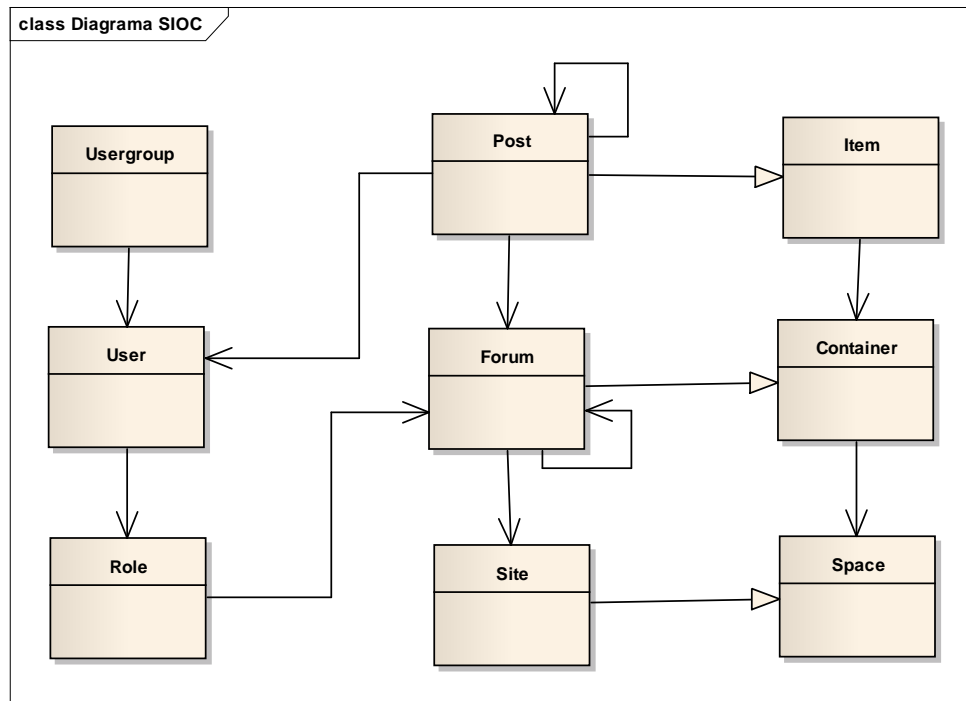


Figura 6. Modelo SIOC simplificado.

El problema de SIOC es que omite los aspectos relacionados con las “Políticas”, así como omite varios otros subaspectos importantes, tales como los permisos, varias otras interacciones que se establecen dentro de una comunidad, que no están relacionadas necesariamente con la publicación de un mensaje, y en general, su lenguaje está muy orientado a determinadas soluciones de software.

2.4.3 SIOC-SNA-DM

El modelo SIOC-SNA-DM [96] [97] es una extensión del modelo SIOC. Tiene como objetivo incorporar aspectos sociales que caracterizan a una comunidad virtual, y que no se encuentran presentes en el modelo original. En la Figura 7, es posible ver que los conceptos de dicho modelo se encuentran divididos en 3 grupos: *Personas*, *Propósito* y *Políticas*.

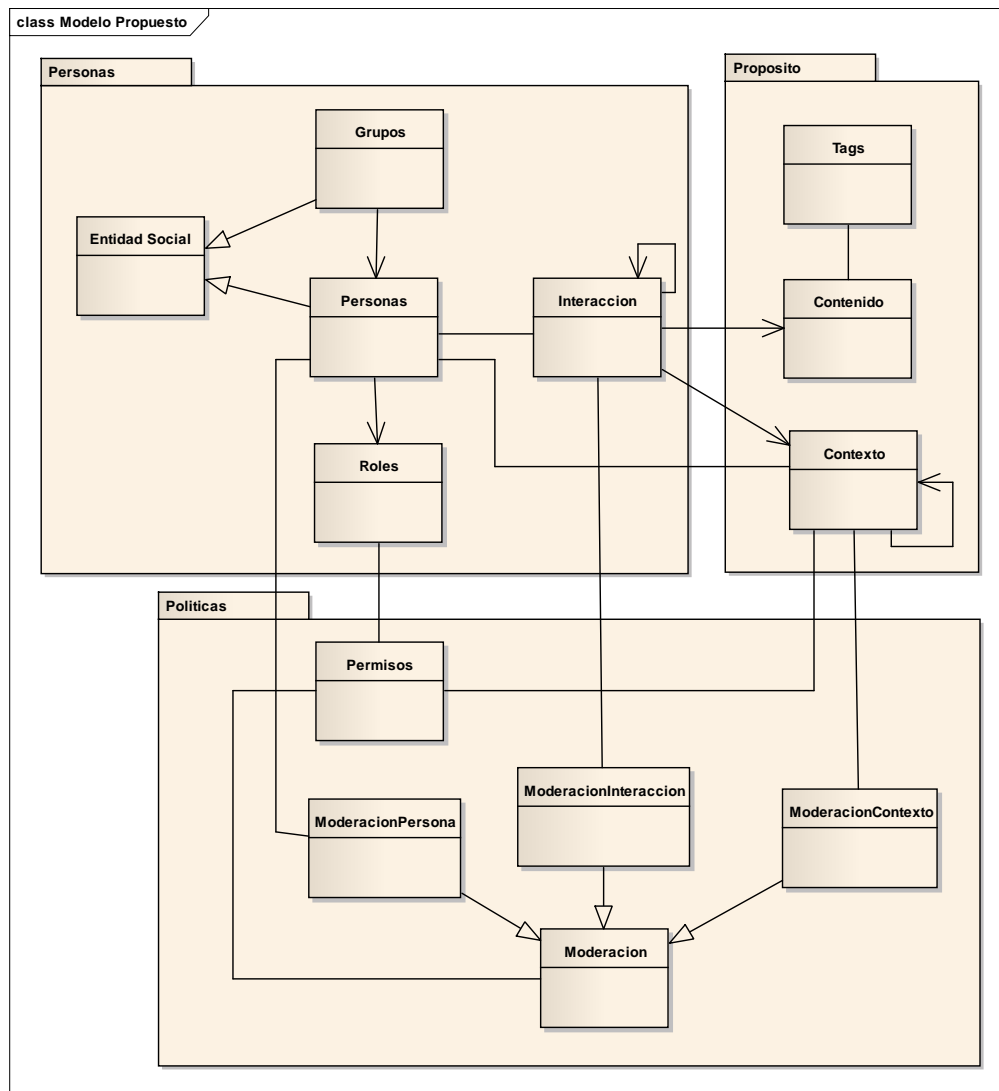


Figura 7. Modelo SIOC-SNA-DM simplificado.

A continuación se presenta una explicación de cada uno de los conceptos:

- *Personas*
 - *Personas*: Corresponde a los miembros de la comunidad. Es el equivalente de *UserAccount* en SIOC (el cual es escrito comúnmente sólo como *User* en los diagramas simplificados) o *People* en FOAF.
 - *Grupos*: Corresponden a agrupaciones de miembros al interior de la comunidad. Es el equivalente de *UserGroup* en SIOC o *Group* en FOAF.

- *Roles*: Corresponden a los roles que cumplen los miembros de la comunidad al interior de ésta (por ejemplo ser ‘Administrador’). Es el equivalente de *Role* en SIOC.
- *Entidad Social*: Modela el hecho de que tanto las Personas como los Grupos son entidades sociales dentro de la comunidad. Es el equivalente de *Agent* en FOAF. Es importante señalar que en ciertos escenarios, las interacciones podrían existir entre Entidades Sociales, en vez de existir entre Personas.
- *Interacción*: Modela una interacción social entre dos o más personas al interior de la comunidad. Una interacción puede ser: responder un mensaje de un usuario, agregar un comentario a un mensaje escrito por otro miembro de la comunidad, enviar un mensaje privado a un miembro de la comunidad, o incluso ver el mensaje publicado por un miembro. La diferencia entre las diferentes interacciones radica en la fuerza del lazo social que se establece. Por ejemplo, visualizar el contenido publicado por un usuario crea un lazo mucho más débil que interactuar con un miembro a través de mensajes privados. En SIOC los conceptos de *Ítem* y *Post* modelan algo similar, pero de forma mucho más limitada.

- *Propósito*

- *Contexto*: Representa el espacio social bajo el cual ocurren las interacciones sociales. Por ejemplo, en un foro de discusión las interacciones ocurren al interior de un tema específico, el cual es parte de un subforo, el cual a su vez es parte del foro completo. Un contexto puede estar dividido en otros contextos. En SIOC sus equivalentes son *Container* y *Space*.
- *Contenido*: Representa el contenido de las interacciones sociales entre los miembros de la comunidad. Es importante destacar que se encuentra bajo Propósito y no Personas, dado que dicho contenido establece la naturaleza de la interacción social, y por tanto, su alineamiento al propósito de la comunidad. En SIOC es parte de *Item*. Una Interacción podría no tener Contenido.
- *Tags*: Modela el conjunto de palabras o frases que describen el contenido de una interacción social.

- *Políticas*

- *Permisos*: Representan los permisos asociados a los roles existentes en la comunidad. Permite establecer además los niveles de permisividad sobre los Contextos existentes.
- *Moderación*: Representa una acción orientada a guiar a la comunidad. Puede ser de tres tipos diferentes: sobre las Personas, sobre las Interacciones o sobre un Contexto existente.

- *Moderación Persona*: Representa una acción de moderación sobre los miembros de la comunidad, por ejemplo, el bloqueo de una persona para interactuar con otros.
- *Moderación Interacción*: Representa una acción de moderación sobre las interacciones al interior de la comunidad, por ejemplo, la eliminación o corrección de un mensaje escrito por un miembro de la comunidad.
- *Moderación Contexto*: Representa una acción de moderación sobre un contexto al interior de la comunidad, por ejemplo, el cierre de un foro de discusión.

Una de las grandes ventajas de este modelo es que permite representar diversas situaciones con un único modelo. Por ejemplo, es posible modelar desde una comunidad que utiliza un foro de discusión, hasta una comunidad que utiliza un sistema de red social. Esto permitirá poder utilizar las mismas técnicas de evaluación para las diferentes comunidades, de tal forma que incluso se podrían comparar los resultados obtenidos entre ellas.

Capítulo 3 – Estrategia de Evaluación Propuesta

El presente trabajo de tesis busca definir una estrategia para evaluar aspectos sociales específicos de una comunidad virtual, específicamente aquellos relacionados con el concepto de sociabilidad (personas, propósito y políticas [5]). El objetivo es mostrar cómo capturar la información relevante, cómo procesarla y representarla a través de indicadores globales. La información es capturada desde el sistema de software utilizado por los miembros de la comunidad, para luego ser almacenada en una base de datos común, basada en el modelo SIOC-SNA-DM, desde la cual se realiza el proceso de análisis. Los indicadores obtenidos pueden luego ser analizados a través del tiempo, de tal forma que provean de información valiosa a los administradores de una comunidad virtual, quienes pueden utilizarlos en los procesos de toma de decisiones relacionadas con la comunidad. A continuación se detalla la estrategia de evaluación propuesta, incluyendo una descripción de los aspectos sociales a evaluar.

3.1 Aspectos Sociales a Evaluar

El presente trabajo se basa en el análisis de los aspectos sociales relacionados con el concepto de sociabilidad; específicamente con personas, propósito y políticas [5], que se describen a continuación:

- i. *Personas* [5][46][45]: Las personas son una parte fundamental de una comunidad virtual, pues son ellas las que la conforman. Sin personas no existen las comunidades. Las discusiones, la generación de nuevas ideas, y el constante cambio de los contenidos son los que distinguen a una comunidad de las páginas Web tradicionales. Una persona es mucho más que un usuario de un sistema. Una persona es una entidad social en sí misma: tiene sus propios objetivos y metas. Las personas son las que se relacionan entre sí para conformar una nueva estructura social. Para un mejor entendimiento, la dimensión personas se ha dividido en los siguientes indicadores, que son los usados por los administradores para entender lo que está pasando con la comunidad desde esa perspectiva:
 - a. *Niveles de Participación* [98][47]: En una comunidad virtual, no todos sus miembros participan de la misma forma. Esto se debe a la voluntariedad de la participación, la cual provoca una situación en constante evolución, en la cual los mismos miembros van cambiando su forma de participar a lo largo del tiempo. Por ejemplo, es posible encontrar moderadores (quienes guían las discusiones), profesionales (quienes dan opiniones) y observadores (quienes silenciosamente observan el funcionamiento de la comunidad). Por lo anterior, es esperable que existan miembros que participen más y otros que participen menos, o incluso sólo observen. Específicamente se espera, en la mayoría de los casos, que alrededor del 10% al 15% de los miembros de la comunidad tenga un nivel de participación alto, entre un 15% al 20% sean miembros activos, mientras que el resto posea un nivel

bajo o nulo de participación [2]. En el presente trabajo se propone medir los niveles de participación a través del uso de técnicas de análisis de redes sociales, utilizando los logs de actividad y la base de datos del sistema de software utilizado por la comunidad, y compararlos con los niveles de participación esperados.

- b. *Identidad* [47][99]: Los miembros de una comunidad, a través del tiempo van generando un lazo “invisible” que los une, y que hace que sientan que pertenecen a una estructura social en común. Este sentido de identidad es lo que permite que los miembros de una comunidad dada, participen en forma voluntaria, y es el elemento que hace que la comunidad sea una estructura viva y que en si misma tenga una identidad propia. En el presente trabajo se propone medir la identidad a través de encuestas a los miembros de la comunidad (para medir la identidad percibida por dichos miembros).
- ii. *Propósito* [5][100][98]: Una comunidad virtual existe por un motivo, por algún propósito. Es ese propósito el que motiva a sus miembros a formar parte de esa comunidad. Sin embargo, sus miembros podrían tener un propósito diferente. Las razones por las cuales una persona pertenece a una comunidad varían. Algunos quieren información o soporte, interactuar con otros, entretenerse, conocer nuevas personas, o escuchar sus propias ideas. Cada participante tiene su propio motivo. Entendiendo los aspectos que motivan a las personas a entrar y retornar a una comunidad, es posible tomar decisiones técnicas y sociales adecuadas. La comunidad debe generar un balance entre estos propósitos, de tal forma que cada miembro pueda cumplir sus propios objetivos, a partir del objetivo común que tiene la comunidad. Para un mejor entendimiento, se ha dividido esta dimensión en los siguientes indicadores:
 - a. *Cumplimiento*: El fenómeno más esperable dentro de una comunidad es que todos sus miembros apoyen un propósito común [2]. De esta forma, todas las interacciones y esfuerzos existentes son para apoyar una o varias ideas comunes. En el presente trabajo se propone medir el cumplimiento utilizando encuestas a los administradores (para definir los objetivos de la comunidad), en conjunto con técnicas de minería de texto, para analizar lo que efectivamente dicen las personas en la comunidad, y si existe relación con los objetivos definidos por los administradores.
 - b. *Claridad*: Si bien una comunidad puede tener definido un objetivo común que mueve a sus miembros a trabajar en forma conjunta, sus miembros pueden percibir este objetivo común de diferentes formas. Mientras para algunos puede ser claro, para otros puede no serlo necesariamente, especialmente para aquellos miembros nuevos [5]. En el presente trabajo se propone medir la claridad a través de encuestas a los miembros de la comunidad (para medir la claridad percibida por dichos miembros).
 - iii. *Políticas* [98][47][99]: Uno de los aspectos fundamentales dentro de una comunidad es la relacionada con las labores de administración. Las comunidades necesitan políticas para dirigir su comportamiento. Específicamente, las políticas son necesarias para, por

ejemplo, establecer los requerimientos para ingresar a una comunidad, definir el estilo de comunicación entre los participantes, las conductas aceptadas, entre otras. Los moderadores deben guiar a la comunidad para que pueda desarrollarse a través del tiempo, a través de tareas específicas que contribuyan a que la comunidad alcance sus objetivos. Sin embargo, el nivel de políticas existente puede afectar directamente la forma en la cual participan los miembros de una comunidad, al romper el balance con los otros aspectos sociales, por ejemplo, al utilizar políticas muy restrictivas o permisivas. Específicamente, en el presente trabajo se medirá el siguiente indicador:

- a. *Nivel de Moderación*: Se refiere a los esfuerzos necesarios dentro de la comunidad, hechos por los moderadores, que permiten mantener su funcionamiento. En el presente trabajo se propone medir los esfuerzos de moderación a través del análisis de los logs de actividad y de la base de datos del sistema de software utilizado por la comunidad.

Un aspecto fundamental es el hecho de que los aspectos sociales señalados anteriormente, se interrelacionan. Esto provoca que, en la búsqueda del apoyo de uno de esos factores, se esté afectando inevitablemente a los otros dos. Por ejemplo, en una comunidad virtual, que trate sobre cierto temas específicos, como por ejemplo, confección de artículos electrónicos. En la medida que se realizan acciones para aumentar la permisividad, el propósito podría divergir rápidamente.

La situación anterior se refleja en la Figura 8. Los 3 aspectos sociales se relacionan entre sí, ocasionado que los cambios provocados en cualquiera de ellos, pueda tener un efecto en los otros dos, los cuales, podrían ser positivos o negativos, dependiendo de la naturaleza del cambio, y el tipo de comunidad sobre el cual se efectúa.



Figura 8. Interrelación de aspectos sociales que posee los sistemas de apoyo a comunidades virtuales.

Las interrelaciones existentes entre los aspectos sociales señalados, resaltan el hecho de la necesidad de incluir todos estos aspectos en cualquier estrategia de evaluación relacionada con apoyo computacional a comunidades virtuales. Sólo de esa forma será posible comprender la naturaleza del comportamiento de las comunidades. La omisión del estudio de algunos de estos

aspectos, podría omitir información importante relacionada con la comunidad virtual, no solamente información ligada a dicho aspecto en particular, sino además información sobre las relaciones e impactos con los otros aspectos.

Es importante destacar que la influencia de un aspecto sobre otro, puede generar un efecto positivo o negativo que depende, entre otras cosas, de la naturaleza de la comunidad. Por ejemplo:

- Si el moderador de la comunidad espera aumentar los niveles de participación de los miembros de la comunidad, a través de un mecanismo de libre participación, en el cual cualquiera puede escribir lo que quiera en el foro de discusión que utiliza la comunidad. Si el moderador omite el hecho de la existencia del propósito, podría suceder que efectivamente en un comienzo exista mayores niveles participación, pero absolutamente fuera del propósito de la comunidad, ocasionado ideas divergentes. Dicha situación puede provocar el efecto contrario a largo plazo, que las personas dejen de participar, puesto que la comunidad ya no les ofrece un espacio para lograr sus objetivos, dado que el propósito de la comunidad cambia con el tiempo, de acuerdo a cómo se comporten sus miembros.
- Si el moderador de la comunidad quiere aumentar los niveles de moderación de la comunidad, con el objetivo de que converja el propósito, es decir, se hablen mayoritariamente de los temas relacionados con la comunidad, y no de otros, el efecto inmediato que podría provocar es la disminución en la participación. Sin embargo, podría provocar el efecto contrario, puesto que, dado el propósito de la comunidad, muchos más personas se motivarían a participar, dado lo específico de la comunidad.

Lo anterior muestra que es importante conocer estas interrelaciones, puesto que determinarán comportamientos que quizás no sean necesariamente los esperados en una comunidad, y que por tanto, es necesario considerarlos en una etapa temprana.

3.2 Descripción de la Estrategia Propuesta

En la Figura 9 es posible ver un esquema general de la estrategia propuesta. Este esquema está orientado a comunidades virtuales que utilizan un sistema de software para mediar las interacciones entre sus miembros. Se debe tener acceso a la base de datos de dicho sistema, para poder ejecutar el proceso de “Extracción de Datos”, el cual obtiene los datos de la comunidad, los almacena en una estructura común (correspondiente a la “Base de Datos Social”, la cual se basa en el modelo SIOC-SNA-DM).

Posteriormente se ejecutan los procesos de “Preparación de Datos”, “Aplicación de Técnicas de Evaluación”, “Evaluación de Resultados” y “Generación Reporte”. El reporte final es elaborado para que lo visualicen los administradores de la comunidad, dado que son ellos quienes conocen la historia de la comunidad, y le dan un sentido y una interpretación a los resultados obtenidos. Además son ellos los que deciden las acciones a seguir, de acuerdo a lo que observen en los resultados.

Los Administradores de la comunidad deben colaborar en ciertas acciones dentro del proceso de evaluación, principalmente en evaluar los resultados obtenidos, o en entregar información que permita realizar la evaluación. Debido a ese hecho el proceso no es completamente automático.

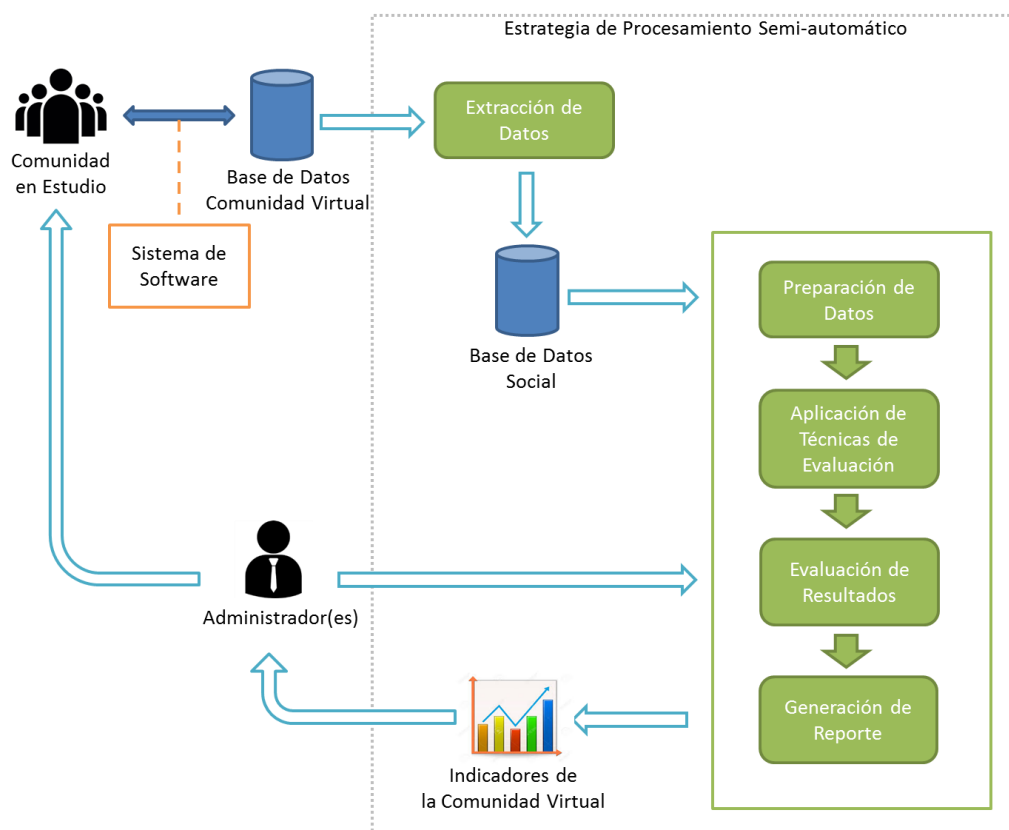


Figura 9. Esquema general de la estrategia propuesta.

3.3 Visualización de los Resultados

Muchas veces no basta con saber que un determinado fenómeno sucede, sino que además es necesario cuantificarlo. Este hecho toma mayor atención en comunidades virtuales, en las cuales no basta con la percepción de que “algo” sucede, es necesario contar con herramientas que permitan a los administradores tener información de lo que realmente sucede. Por ejemplo, un administrador podría notar que la participación baja al interior de la comunidad, pero mientras no tenga números claros de lo que sucede realmente, es sólo una percepción.

El presente trabajo se enfoca en utilizar indicadores, los cuales son medidos y posteriormente graficados, de tal forma que el administrador pueda ver cómo ha ido cambiando su valor a través del tiempo. Posteriormente, utilizando el conocimiento del administrador, se podría establecer si

un determinado cambio en el comportamiento de un aspecto social se debe o no a alguna intervención o mejora realizada por el propio administrador (Figura 10).



Figura 10. Visualización de la evolución temporal de un aspecto social en forma de gráfica.

El hecho de que los 3 aspectos sociales se encuentren relacionados, plantea una característica especial: es posible evaluarlos y analizar sus resultados en conjunto. Para ello, es posible obtener un “promedio” de dichos aspectos, de tal forma de determinar cómo se encuentra la comunidad, desde una perspectiva general (ubicando dicho promedio al interior del modelo).

La Figura 11 muestra tres ejemplos. En el primero el “promedio” de los tres aspectos sociales se encuentra justo en el centro, lo cual indica que hay un balance entre los 3 aspectos sociales al interior de la comunidad. En el segundo ejemplo el “promedio” se encuentra cercano al aspecto “Personas”, lo cual indica que la comunidad se encuentra centrada en dicho aspecto, más que los otros dos. Ese podría ser el caso de una comunidad que utiliza un sistema de redes sociales, y en el cual los miembros establecen relaciones entre ellos, pero no hay un propósito definido ni tampoco es necesario moderar. En el tercer ejemplo el “promedio” se encuentra cercano tanto a las “Personas” como a las “Políticas,” pero absolutamente lejano al “Propósito”, es decir en la comunidad existen relaciones fuertes, existen esfuerzos de los administradores en manejar la comunidad, pero no hay un propósito definido.

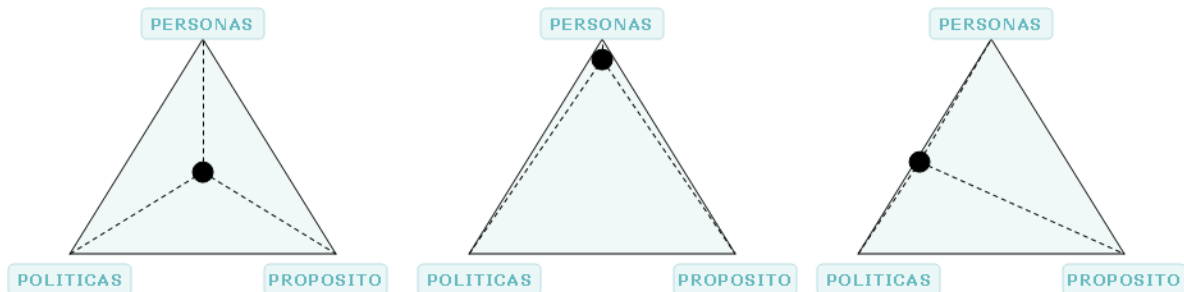


Figura 11. Utilización del modelo para evaluar una comunidad virtual.

La situación anterior le permite obtener una visión clara al administrador de la comunidad sobre qué sucede con la comunidad. Si a eso sumamos el hecho de que dichos valores cambian sobre el tiempo, el administrador podría comprender ciertos fenómenos desde la perspectiva social, de tal forma que pueda tomar decisiones con mucha más información. Adicionalmente, el administrador contará con una herramienta que le permita saber que aspecto social presenta falencias, y de acuerdo a los objetivos de la comunidad, pueda determinar si requiere apoyarlo de mejor forma, pudiendo posteriormente volver a evaluar.

Capítulo 4 – Los Aspectos Sociales en Detalle

A continuación se explican en detalle cada uno de los aspectos sociales abordados en este trabajo de tesis: *personas*, *propósito* y *políticas*.

4.1. Evaluación de Personas

Todas las comunidades virtuales son redes sociales, y como tales un elemento central en ellas son las relaciones sociales que se establecen entre sus miembros. Estas relaciones son la base sobre la cual se crea el sentimiento compartido de comunidad, y sobre la cual se crea una identidad común entre sus miembros. Para que puedan existir estas relaciones sociales, es necesario que sus miembros interactúen en forma sostenida a través del tiempo, de tal forma que se formen lazos que los unan y no sean sólo personas aisladas que forman un grupo sin identidad.

Por lo tanto, la interacción entre los miembros de la comunidad es vital. No es solamente que participen en la comunidad, sino que dicha participación genere algún grado de interacción con otros miembros. Es por ello que, desde el punto de vista social, es más importante una persona que participa y genera que otros miembros participen, que aquellos miembros que, participando mucho, no generan otras interacciones.

Por ejemplo, para ilustrar la idea anterior, si un miembro de la comunidad virtual afirma que “el cielo es rojo” al interior de la comunidad, y esto provoca que la mayoría de los miembros participen, entonces, desde el punto de vista de “Personas”, dicho miembro es importante. Sin embargo, como ya se ha mencionado, puede que sea algo negativo que se realicen muchas interacciones sobre un tema no perteneciente a la comunidad, pero desde este aspecto social, sería bueno que ocurra.

Para poder evaluar a las “Personas”, muchos de los algoritmos que provee SNA resultan idóneos, puesto que parte de la importancia de una persona está dada por la importancia que tiene al interior de la red social subyacente. En la medida que la participación de una persona en una comunidad virtual produce interacción con otros miembros, es que dicha participación se considerará mucho más valiosa que aquellos miembros que participan en forma aislada en la comunidad. Los algoritmos de SNA pueden medir precisamente eso, cuán importante es una persona respecto de la red en la que participa.

4.1.1 Niveles de Participación

En una comunidad, tal como lo menciona Wenger [39], no todos los miembros participan de la misma forma, ni con la misma intensidad. La participación de un miembro de una Comunidad Virtual está orientada a satisfacer sus propias necesidades y de acuerdo a las posibilidades que le ofrece el sistema de software que media la interacción con los otros miembros. Dado que la

participación es principalmente voluntaria en este tipo de estructuras sociales, los niveles de participación de un miembro no son siempre los mismos, pudiendo variar durante el tiempo, así como también puede ir variando los miembros que participan en la comunidad. Esto plantea el primer desafío: poder medir la participación que existe en una comunidad virtual. Sin embargo, no todas las participaciones son iguales. Existirán miembros que participen más que otros. Wenger [39] señala por lo menos 3 tipos de niveles de participación: Núcleo, Activos y Periferia.

Un sistema de software debe permitir diferentes niveles de participación al interior de la comunidad, de tal forma que sus miembros utilicen el medio que sientan más adecuado de acuerdo a su nivel de interés con la comunidad. De esta forma, algunos miembros decidirán solo mirar, otros opinar, y otros generar contenido, entre otros. Por lo tanto, no es necesario que todos los miembros de la comunidad participen de la misma forma, es importante que exista esta diversidad en la participación, dado que es esta dinámica la que permite mantener viva la comunidad.

Métricas de Centralidad en SNA

En redes sociales en línea se plantea frecuentemente que un número reducido de miembros son capaces de generar gran parte de la participación de la red social completa [101]. Este mismo hecho es algo conocido en el mundo de las comunidades virtuales. Wenger [39] plantea que aproximadamente del 10 al 15% de los miembros de una Comunidad de Práctica pertenecen al núcleo de la comunidad, y otro 10 al 15% pertenece a los miembros activos. Lo anterior señala que entre un 20 al 30% de la comunidad son los miembros más importantes de la comunidad, siendo el resto (del 70 al 80%) parte de la periferia de los miembros, es decir, corresponden a miembros que sólo participan en forma ocasional. A continuación se enumeran las técnicas de centralidad que son utilizadas:

Degree Centrality

El grado de un vértice $d(v)$ es una medida que cuenta el número de aristas incidentes a v .

$$d(v) = \sum_j (a_{ij})$$

Mediante la evaluación de la centralidad de esta forma, se compara la conectividad de los vértices sin medir ni mostrar qué tan bien posicionado se encuentran los vértices dentro del grafo.

Closeness Centrality

Es una medida que evalúa cuan cercano se encuentra un vértice a los otros vértices de un grafo. Si $\text{dist}(s,v)$ es la distancia más corta entre los vértices s y v , la centralidad de un vértice v se calcula de la siguiente forma:

$$c(v) = \sum_{s \in G \setminus \{v\}} \frac{1}{\text{dist}(s, v)}$$

Betweenness Centrality

La idea principal de esta métrica es que los vértices que se encuentran en el camino geodésico (aquel camino de longitud menor entre todos los posibles caminos que unen dos nodos en un grafo) de muchos otros vértices poseerán un gran control sobre el flujo de información debido a que ellos se encuentran “entre” (*between*) los otros vértices. La centralidad de un vértice v para esta métrica se calcula como:

$$b(v) = \sum_{s \in V \setminus \{v\}} \sum_{t \in V \setminus \{s, v\}} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

En donde σ_{st} es el número de caminos geodésicos entre los vértices s y t . El valor $\sigma_{st}(v)$ corresponde al número de caminos geodésicos entre s y t , en los cuales v pertenece a dichos caminos.

Hubs y Authority

Son dos indicadores que son obtenidos a partir del algoritmo HITS, el cual es utilizado para valorar, y de paso clasificar, la importancia de una página web. Authority valora cuan buena es una página como recurso de información. Hub dice cuan buena es la información que se consigue siguiendo los enlaces que tiene a otras páginas. Ambas métricas pueden ser utilizadas para calcular la centralidad de vértices al interior de un grafo.

Pagerank

También corresponde a un algoritmo que se utiliza originalmente para medir la importancia de páginas web. La centralidad $P(j)$ es una variante de la centralidad basada en vectores propios y puede ser determinada por la siguiente ecuación:

$$P(j) = \frac{(1-d)}{n} + d \sum_{i \in B_j} \frac{P(i)}{|F_i|}$$

En donde d es probabilidad de saltar en forma ocasional de una página a otra, n corresponde al total de nodos (páginas en este caso), F_i el conjunto de vértices (páginas) a los cuales el vértice i apunta, B_i el conjunto de vértices (páginas) que apuntan a i . Dado que es una función recursiva, se evalúa de forma iterativa hasta que el valor de $P(j)$ converja.

4.1.2 Proceso de Evaluación

El proceso de evaluación del indicador “Niveles de Participación” se presenta a continuación:

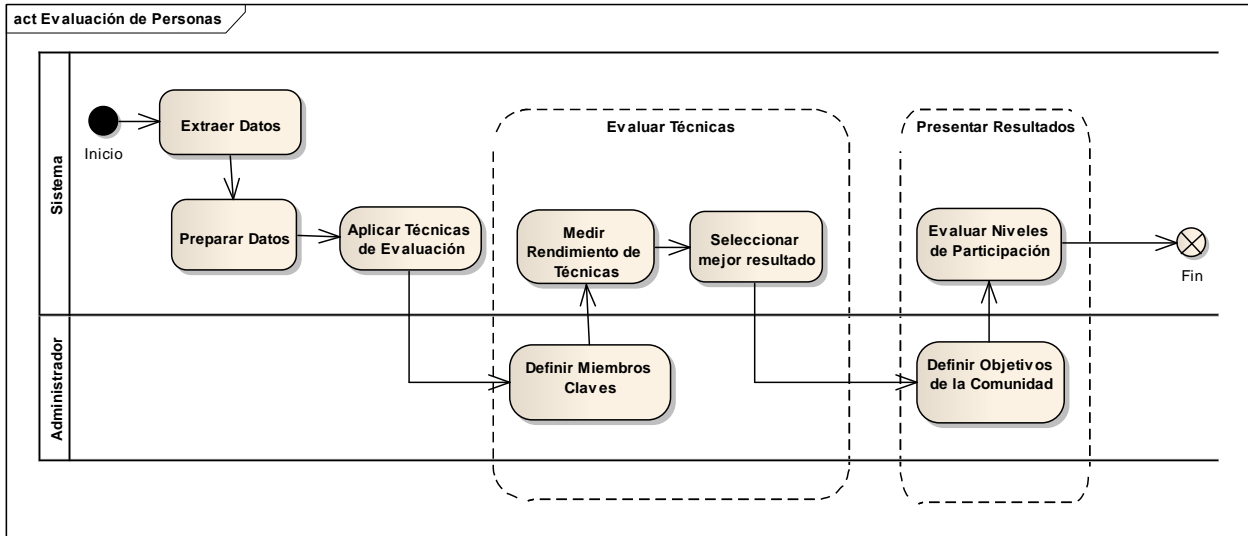


Figura 12. Proceso de Evaluación de Niveles de Participación.

- Extraer datos:* A partir del modelo de aspectos sociales (SIOC-SNA-DM), es necesario extraer la información relevante para realizar el estudio. En particular, es necesario obtener los datos de *Personas*, *Contexto* e *Interacciones*.
- Preparar datos:* Consiste principalmente en generar las redes sociales subyacentes a partir de las interacciones entre los miembros de la comunidad virtual. Para realizar dicha tarea, existen por lo menos 4 enfoques (Ver Figura 13):

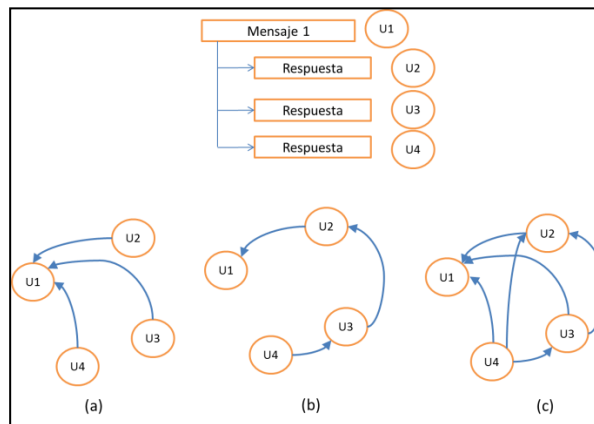


Figura 13. Formas de representar una red social a partir de las interacciones entre los miembros: (a) creador, (b) reply_prev, (c) reply_all.

- a. *creator*: Cuando un usuario A escribe un mensaje en una discusión X, dicho mensaje es una respuesta sólo al que inició la discusión.
 - b. *reply_prev*: Cuando un usuario A escribe un mensaje en una discusión X, dicho mensaje es una respuesta sólo al último mensaje.
 - c. *reply_all*: Cuando un usuario A escribe un mensaje en una discusión X, dicho mensaje es una respuesta a todos los mensajes anteriores.
 - d. *reply_all30*: Cuando un usuario A escribe un mensaje en una discusión X, dicho mensaje es una respuesta sólo a los mensajes anteriores que hayan sido escritos con una antigüedad de a lo más 30 días.
- c) *Aplicar las técnicas de evaluación*: Se calculan las centralidades de todos los miembros de la comunidad, para todo el periodo de actividad de la comunidad virtual, utilizando las diferentes métricas de centralidad (*betweenness centrality*, *closeness centrality*, *degree*, *pagerank* y *hubs/authority*), y las diferentes representaciones de redes sociales mencionadas anteriormente. Con dicha información, se calculan los niveles de participación de la siguiente forma:
- a. *Miembros activos*: aquellos miembros que hayan interactuado durante el periodo en estudio, y que su índice de centralidad sea distinta de 0.
 - b. *Miembros de la periferia*: Aquellos miembros vigentes que no interactuaron en el periodo dado o que su índice de centralidad es 0.
 - c. *Miembros del núcleo*: Aquellos miembros que tienen el índice de centralidad más alto, basados en la curva de distribución acumulada del índice de centralidad calculado.
- d) *Evaluar las técnicas*: Utilizando la información provista por los administradores, se compara con los datos obtenidos por las diferentes métricas de centralidad, y se determina cuál técnica y sobre cuál tipo de red social genera mejores resultados. Para ello se realizan los siguientes pasos:
- a. *Definir los miembros claves*: Es necesario que los administradores de la comunidad virtual definan, para un periodo dado, cuáles fueron los miembros claves o usuarios más importantes para la comunidad.
 - b. *Medir rendimiento de técnicas*: Para cada resultado obtenido de la aplicación de las técnicas de evaluación, y con la información entregada por los administradores, se calculan métricas que indiquen cuál técnica ofrece mejor rendimiento en la determinación de miembros perteneciente al núcleo de la comunidad virtual. Para lo anterior, se utiliza la métrica F_1 , la cual se expresa como:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

En donde:

$$precision = \frac{|\{miembros\ claves\ definidos\ por\ administradores\} \cap \{núcleo\}|}{|\{núcleo\}|}$$

$$recall = \frac{|\{miembros\ claves\ definidos\ por\ administradores\} \cap \{núcleo\}|}{|\{miembros\ claves\ definidos\ por\ administradores\}|}$$

- c. *Seleccionar mejor resultado:* se selecciona la técnica que tenga mejor índice F_1 .
- e) *Presentar resultados:*
 - a. *Definir objetivos de la comunidad:* Si bien en la literatura se plantea cual fracción de miembros debiera poseer ciertos niveles de participación, se deja una instancia para que los administradores de la comunidad puedan definir valores diferentes, de acuerdo a los objetivos que ellos planteen para la comunidad misma. Se toman como referencia los valores dados por Wenger, esto es: entre 10 y 15% el tamaño del núcleo (core), y 15% a 20% el tamaño de los miembros activos.
 - b. *Evaluar los niveles de participación:* Se calculan los niveles de participación de los miembros de la comunidad, para todo el periodo en el cual se encuentra activa la comunidad virtual. Se generan semáforos para mostrar si el nivel de participación real difiere de los niveles de participación definidos por los administradores.

4.1.3 Identidad

Una característica fundamental de los sistemas de software utilizados en comunidades virtuales, es que sus usuarios son personas, y como tales, tienen ciertas expectativas y percepciones acerca del sistema de software que utilizan. Evaluar únicamente las acciones que ellos realizan, sin considerar cómo ellos perciben su pertenencia a una comunidad virtual, sería evaluar sólo una parte de los aspectos sociales involucrados.

El concepto de identidad se encuentra fuertemente ligado al concepto de sentido de comunidad [102], y como tal, su foco está en la percepción de los miembros acerca de los aspectos de la comunidad a la que pertenecen.

Puddifoot [103], basado en las teorías de Sentido Psicológico de Comunidad, Cohesión Social y Satisfacción Comunitaria, define un conjunto de dimensiones que caracterizan el sentido de identidad comunitaria, tanto para comunidades territoriales, como también para comunidades basadas en relaciones sociales o culturales. Entre las dimensiones que se destacan se encuentran: (a) *locus*: el cual hace referencia a la percepción que tienen los miembros acerca de los límites de la comunidad, (b) *diferencia*: el cual se refiere a los aspectos distintivos percibidos de la

comunidad, (c) *identificación*: el cual se refiere a la sensación percibida de afiliación y pertenencia, (d) *orientación*: el cual expresa la orientación personal de un miembro a la comunidad desde los puntos de vistas de grado de inversión personal en la comunidad, atracción a la comunidad y sentido de seguridad emocional, entre otros.

Si bien existen escalas de evaluación que incluyen el concepto de identidad, tales como SCI [104] (*Sense of Community Index*), o el PSCS [105] (*Psychological Sense of Community Scale*), no son completamente aplicables, dado que fueron creadas específicamente para comunidades territoriales. Existen varias adaptaciones creadas especialmente para comunidades virtuales, tales como SOCV [106] (*Sense of Virtual Community*), o mejoras de versiones anteriores que incluyen más tipos de comunidades, tales como SCI-2 [107] (*Sense of Community Index 2*).

Basados en las dimensiones y escalas anteriores, se definieron un conjunto de preguntas para medir los niveles de identidad percibida por los miembros de las comunidades virtuales en estudio. Se utiliza una escala Likert para responder cada pregunta, la cual está compuesta por 5 valores posibles, desde Muy de Acuerdo a Muy en Desacuerdo.

Aspecto a Evaluar	Pregunta
Locus	¿Te sientes parte de la comunidad?
Identificación	¿Sientes que hay otros miembros de la comunidad que tienen tus mismos intereses?
Orientación	¿Participas en forma frecuente en la comunidad X?
Diferencia	¿Es importante para ti ser un miembro de la comunidad X?

Con lo anterior, la encuesta queda de esta forma:

Preguntas	
	muy en desacuerdo muy de acuerdo
1. ¿Te sientes parte de la comunidad X?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1 2 3 4 5
2. ¿Sientes que hay otros miembros de la comunidad que tienen tus mismos intereses?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1 2 3 4 5
3. ¿Participas en forma frecuente en la comunidad X?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1 2 3 4 5
4. ¿Es importante para ti ser un miembro de la comunidad X?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1 2 3 4 5

4.2 Evaluación del Propósito

Como se ha mencionado anteriormente, en una comunidad virtual es importante que los miembros establezcan relaciones sociales, pero con el objetivo de cumplir un determinado propósito. Sin un propósito, la comunidad sólo sería un grupo de personas interactuando. Dicho propósito es lo que los une, y por lo tanto, es necesario analizar en el tiempo si efectivamente sigue manteniéndose o va cambiando.

Desde el punto de vista del propósito un hecho fundamental es el contenido de las interacciones sociales que se forman, más que las relaciones sociales en sí. Por ejemplo, sería mucho más importante, desde este punto de vista, que los miembros de la comunidad hablen de los temas de la comunidad, aunque generen poca participación entre ellos, a que hablen de temas no relacionados, pero generando mucha participación.

Para poder medir este aspecto, es necesario utilizar técnicas que analicen el contenido de las interacciones, de tal forma de poder conocer si los aportes, interacciones o información que generan los miembros de la comunidad se encuentran alineados al propósito de ésta.

4.2.1 Cumplimiento del Propósito

En una comunidad es deseable que existan metas comunes que una a los miembros de la comunidad. Los miembros de una comunidad pertenecen a ella porque comparten algún interés, propósito o práctica. Esto último genera que los miembros compartan un lenguaje común, el cual tiene directa relación con la existencia de la comunidad.

Ahora bien, ¿Quién debiera saber de qué debe hablar los miembros de la comunidad? ¿Cómo establecer si lo que habla una comunidad es correcto o no? La respuesta está en los dueños y administradores de la comunidad. Son ellos quienes hacen los esfuerzos para que los miembros se encuentren alineados al propósito compartido que tiene la comunidad. Son ellos quienes habilitan funcionalidades o sistemas de software nuevos para que los miembros de la comunidad puedan interactuar. Por lo tanto, son ellos quienes conocen de qué debiera hablar la comunidad. De esta forma, la participación de ellos en la evaluación del propósito es fundamental.

Sin embargo, a pesar de lo anterior, surgen un par de dudas fundamentales: si los miembros de una comunidad hablan específicamente de los temas que esperan los administradores el análisis termina ahí. Pero ¿qué pasa si no es así? Existen dos posibilidades: (a) que los miembros estén hablando de otros temas, no conocidos por el administrador, los cuales podrían o no estar alineados a los objetivos de la comunidad, o (b) que no estén hablando de ningún tema en particular. Lo anterior implica nuevamente que el uso de la técnica específica de evaluación debe necesariamente complementarse con la participación de los administradores.

Para poder evaluar el cumplimiento del propósito se propone aplicar técnicas de minería de datos [108][109][110][111][112], específicamente utilizando técnicas de minería de datos basadas en modelos de tópicos.

4.2.1.1 Modelos de Tópicos

Un modelo de tópicos puede ser considerado como un modelo probabilístico que relaciona documentos y palabras, con variables que representan los principales tópicos inferidos desde un texto. En este contexto, un documento puede ser considerado como una mezcla de tópicos, pero en diferentes proporciones. El proceso de inferencia de las variables latentes, o tópicos, es la componente clave de este modelo, cuyo principal objetivo es aprender desde el texto de los documentos la distribución de los tópicos.

De esta forma, un modelo de tópicos es capaz de descubrir un conjunto de “tópicos” abstractos que ocurren en una colección de documentos, cada uno caracterizado por un conjunto de palabras con cierta probabilidad de ocurrencia. Los modelos utilizados en el presente trabajo son los siguientes:

Latent Dirichlet Allocation (LDA)

LDA [108][109] es básicamente un modelo Bayesiano en donde los tópicos latentes de los documentos son inferidos a partir de distribuciones de probabilidad estimadas sobre un conjunto de datos de entrenamiento. La idea principal de LDA es que cada tópico es modelado como una distribución de probabilidad sobre un conjunto de palabras que son representadas por el vocabulario V , y cada documento como una función de probabilidad sobre un conjunto de tópicos (T). Estas distribuciones son generadas desde una distribución *Dirichlet multinomial*. La representación del modelo es mostrada en la siguiente figura.

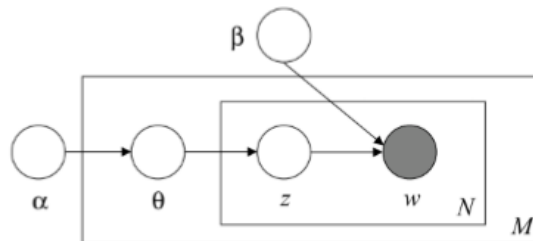


Figura 14. Representación del modelo LDA (Figura obtenida de [113]).

En el modelo, se define que:

- Una palabra es la unidad base de datos, y corresponde a un ítem de un vocabulario indexado por $\{1, \dots, V\}$.
- Un documento es una secuencia de N palabras denotado por $W = (w_1, w_2, \dots, w_N)$, donde w_n es la n -sima palabra de la secuencia.

- Un corpus es una colección de M documentos denotado por $D = (W_1, W_2, \dots, W_M)$.
- z representa un conjunto de N tópicos.
- α es un parámetro sobre la distribución de tópicos por documento.
- β es un parámetro sobre la distribución de palabras por tópico.
- θ_i es la distribución de tópico de un documento i .

Dado los parámetros β y α , el objetivo es determinar la distribución conjunta de una mezcla de tópicos θ , definida como :

$$p(\theta, z, W | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(W_n | z_n, \beta)$$

Pitman-Yor Topic Model (PYTM)

Es un modelo basado en LDA, el cual usa un proceso Pitman-Yor [114] para generar el prior. Dicho proceso es una distribución sobre un espacio de probabilidades. Éste usa tres parámetros: un parámetro de concentración γ , un parámetro de descuento d ($0 < d < 1$) y una distribución base G_0 . El proceso Pitman-Yor es una generalización del proceso Dirichlet, en el cual el parámetro de descuento es considerado como cero [114].

Este modelo se basa en el Chinese Restaurant Process (CRP), el cual corresponde a un proceso en donde n clientes se sientan en un restaurante chino con un número infinito de mesas. Siempre el primer cliente se sienta en la primera mesa. Los siguientes clientes se sientan en una mesa ocupada o en una nueva mesa desocupada. Para PYTM la representación del CRP está compuesta por 4 elementos: un cliente, una mesa, un plato y el restaurante. El cliente es representado por una palabra en un documento. La mesa es representada por una variable latente. El plato es representado por un tipo de palabra. El restaurante es representado por un documento.

En el CRP, dado que si se aumenta el valor de n entonces aumenta la cantidad de mesas ocupadas, muchas con un sólo cliente, es que este proceso resulta útil para capturar la distribución de ley de potencia (*power law*) que siguen las palabras.

La representación del modelo es mostrada en la Figura 15.

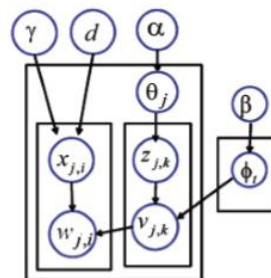


Figura 15. Representación del modelo de tópicos Pitman-Yor (Figura obtenida de [114]).

En el modelo, se define que:

- $x_{j,i}=k$, indica que el i -simo cliente (palabra) se encuentra sentado en la k -sima mesa del restaurant (documento) j .
- $v_{j,k}=v$, indica que el plato (tipo de palabra) v es servido en la k -sima mesa del restaurant (documento) j .
- $w_{j,i}$ corresponde al i -simo cliente (palabra) del restaurante (documento) j .
- $z_{j,k}$ corresponde al t3pico asignado a la k -sima mesa en el restaurante (documento) j .

4.2.1.2 Proceso de Evaluaci3n

El proceso de evaluaci3n del indicador “Cumplimiento” se presenta a continuaci3n:

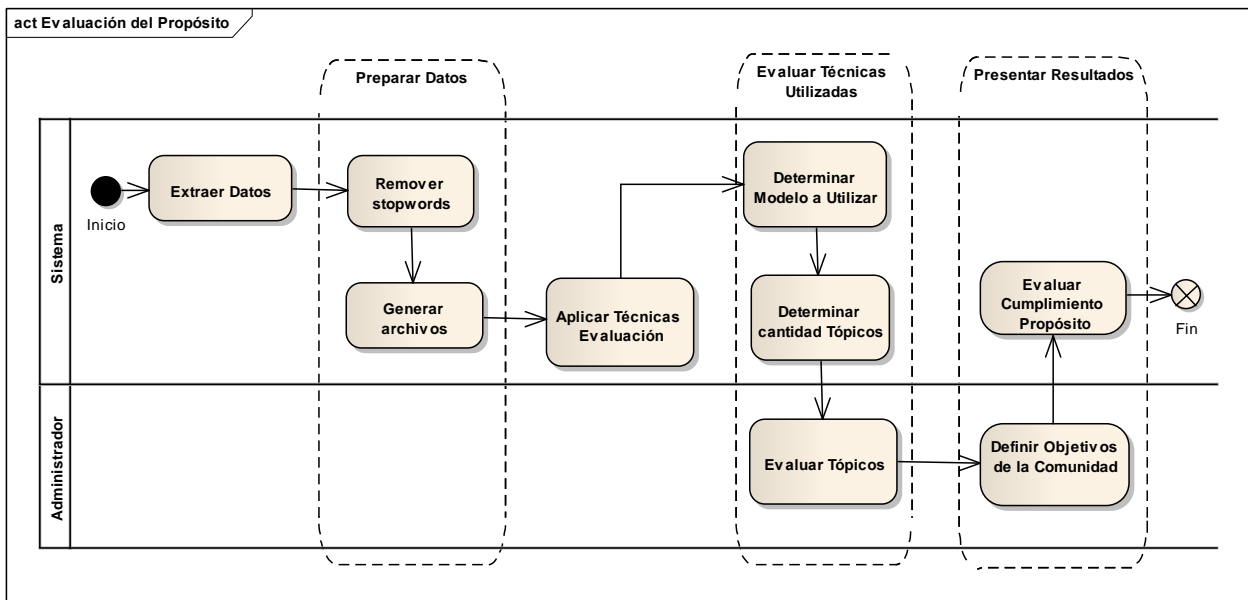


Figura 16. Proceso de Evaluaci3n de Cumplimiento de Prop3sito.

- a) *Extraer datos:* A partir del modelo de aspectos sociales (SIOC-SNA-DM), es necesario extraer la informaci3n relevante para realizar el estudio. En particular, es necesario obtener los datos de *Contenidos*, *Contextos* e *Interacciones*.
- b) *Preparar datos:* Es necesario preparar los datos, de acuerdo al formato utilizado por las t3cnicas de evaluaci3n, y de acuerdo a los procesos de limpieza y transformaciones necesarias para su an3lisis. En particular, es necesario:
 - a. *Remover stopwords:* Muchas de las palabras que se utilizan en un espacio sem3ntico, corresponden a conectores o palabras cortas. Estas palabras son denominadas *stopwords*. Al no contener importancia significativa para ser utilizadas en este trabajo, es necesario removerlas. Para esto, se utiliza una base de 350 *stopwords* en espa3ol.

- b. *Generar archivos*: Es necesario generar archivos en el formato que son leídos por los modelos de tópicos, esto es, un archivo en que cada línea representa un documento, y en el cual en la primera línea se encuentra la cantidad total de documentos.

En la solución final no se utilizan técnicas de lematización ni de stemming, debido a que fue más fácil para los administradores de la comunidad interpretar las palabras originales que las palabras generadas por estas técnicas. Lo anterior se debe a la naturaleza del idioma español, específicamente a los problemas derivados de la polisemia. Por ejemplo, para una de las comunidades la palabra “cuenta” fue transformada a “contar”, la cual no tuvo ningún significado para los administradores de esa comunidad.

- c) *Aplicar las técnicas de evaluación*: Se ejecutan los modelos de tópicos, específicamente LDA y PYTM. Dado que ambas técnicas reciben de parámetro la cantidad de tópicos, la cual es desconocida a priori, se ejecutan los modelos de tópicos con diferentes configuraciones, de tal forma de poder estimar posteriormente la cantidad óptima para cada periodo en estudio.
- d) *Evaluar las técnicas utilizadas previamente*: Es necesario evaluar el rendimiento de los modelos ejecutados para los resultados obtenidos anteriormente. Para ello se utiliza la métrica de perplejidad, la cual corresponde a una medida para comparar modelos de probabilidades. Esta medida captura qué tan bien una muestra es predicha por un modelo. Basada en la entropía de la distribución subyacente, esta medida identifica cuán difícil es para el modelo escoger una palabra en la distribución. Utilizando D como la colección de documentos, M como el número de documentos, N_d como el número de palabras en el documento d , y $p(w_d)$ como la probabilidad asignada por el modelo a una palabra w_n , esta medida es calculada como se muestra a continuación:

$$perplexity(D) = \exp\left\{-\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d}\right\}$$

Un menor valor de perplejidad indica un mejor rendimiento general del modelo. Utilizando esta métrica, se realizan los siguientes pasos:

- a. *Determinar qué modelo utilizar, de acuerdo al que tenga mejor rendimiento con los datos analizados*: Se aplica la métrica a todos los periodos con los diferentes modelos, con las diferentes configuraciones, y se comparan entre sí.
- b. *Determinar cuál es la cantidad óptima de tópicos por periodo*: Para el modelo seleccionado anteriormente, con diferentes configuraciones, se calcula la métrica y luego se determina cuál es la cantidad óptima de tópicos por cada período.
- c. *Evaluar tópicos*: Los tópicos generados por los modelos deben ser revisados por los administradores, para determinar cuáles tópicos corresponden a temáticas que son deseables que se hablen al interior de la comunidad, y cuáles no.

e) *Presentar resultados:*

- a. *Definir objetivos de la comunidad:* Los administradores deben definir los niveles de cumplimiento del propósito que quieren para la comunidad, en términos de qué fracción del total de interacciones entre los miembros de la comunidad debiera pertenecer al propósito de la comunidad.
- b. *Evaluar el cumplimiento del propósito:* Con la evaluación de los administradores, se clasifican las interacciones de las Personas entre aquellas que pertenecen a las temáticas definidas por los administradores y aquellas que no. Se generan semáforos para mostrar el nivel de cumplimiento del propósito para la comunidad virtual en estudio.

4.2.2 Claridad del Propósito

Es una medida cualitativa que está orientada a evaluar qué tan claro es el propósito de la comunidad para sus miembros. En la medida que los miembros tengan claro el propósito, su participación debiera encontrarse más alineada a los objetivos de la comunidad durante el tiempo. Por el contrario, si los miembros tienen poca claridad del propósito de la comunidad, tenderán a participar en temas no relacionados con la comunidad y su propósito.

Mientras que la medida de cumplimiento del propósito puede ser hecha en forma cuantitativa y de forma semi-automática, la medida relacionada con la claridad de la comunidad debe realizarse en forma cualitativa, al igual que la medición de la identidad. Para lo anterior, los miembros de la comunidad virtual en estudio deben contestar las siguientes preguntas:

Preguntas											
	muy en desacuerdo muy de acuerdo										
1. ¿Tienes claro para qué es la comunidad X?	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="text-align: center;">1</td><td style="text-align: center;">2</td><td style="text-align: center;">3</td><td style="text-align: center;">4</td><td style="text-align: center;">5</td></tr></table>						1	2	3	4	5
1	2	3	4	5							
2. ¿Sientes que los miembros de la comunidad tienen claro cuál es el objetivo de la comunidad?	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="text-align: center;">1</td><td style="text-align: center;">2</td><td style="text-align: center;">3</td><td style="text-align: center;">4</td><td style="text-align: center;">5</td></tr></table>						1	2	3	4	5
1	2	3	4	5							

4.3 Evaluación de las Políticas

Los administradores de las comunidades tienen una labor que es fundamental: establecer las bases para que los miembros de la comunidad puedan interactuar adecuadamente y sin problemas. Para ello necesitan frecuentemente efectuar labores de moderación sobre las interacciones de los miembros de la comunidad de tal forma de corregir conductas inapropiadas o que no se encuentre acordes a la comunidad o su propósito.

Si bien el ideal es una comunidad que sea capaz de auto-moderarse, la realidad indica que la acción de los administradores es vital, dado que son ellos los que son capaces de resolver cualquier conflicto al interior de la comunidad, y tomar las medidas adecuadas para prevenir futuros conflictos.

Desde el punto de vista de las políticas, es fundamental que las intervenciones de los administradores sea la menor posible, de tal forma que la comunidad crezca de forma natural. Si bien cada acción de los administradores debiera llevar a un mejor funcionamiento de la comunidad, cada acción hecha por ellos también se verá como una intervención por parte de los miembros de la comunidad. En la medida que los administradores intervienen excesivamente el funcionamiento de la comunidad: (a) los miembros de la comunidad no se sentirán cómodos, (b) significa que el trabajo de mantención de la comunidad es excesivo, y (c) la comunidad se reducirá a un grupo de administradores.

Para poder determinar este indicador es necesario medir la participación de los administradores al interior de la comunidad. Ésta puede ser vista tanto del punto de vista de las interacciones que forman parte de acciones de moderación, como por las labores de moderación propiamente tal, como por ejemplo eliminar un mensaje de un usuario. El proceso de evaluación del indicador “Niveles de Moderación” se presenta a continuación:

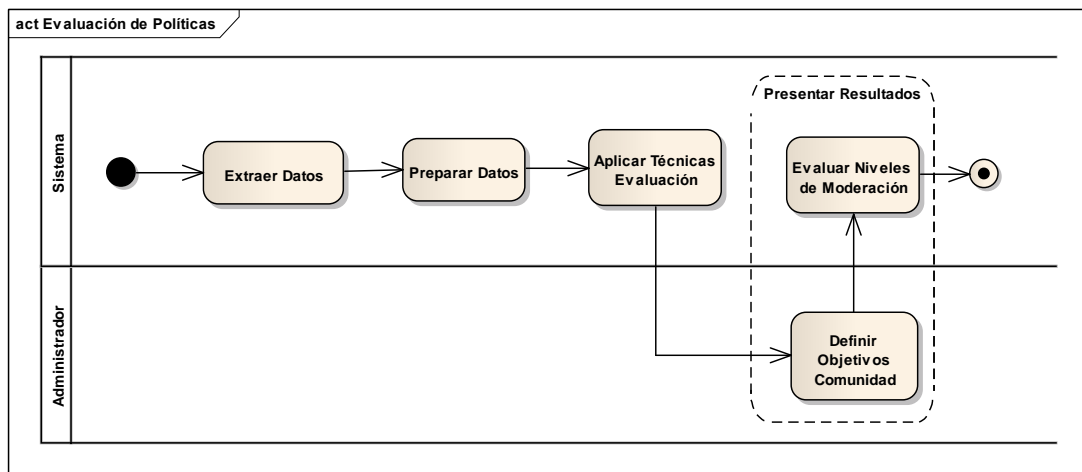


Figura 17. Proceso de Evaluación de Niveles de Moderación.

- a) *Extraer datos:* A partir del modelo de aspectos sociales (SIOC-SNA-DM), es necesario extraer la información relevante para realizar el estudio. En particular, es necesario obtener los datos de *Moderación-Persona*, *Moderacion-Contexto*, *Moderacion-Interacción*, *Interacciones* y *Contextos*.
- b) *Preparar datos:* Consiste principalmente en seleccionar los Contextos en los cuales algún administrador tuvo que intervenir, ya sea interactuando con otros miembros, o realizando alguna labor explícita de moderación.

- c) *Aplicar las técnicas de evaluación:* Se calculan los niveles de moderación de todos los administradores en conjunto, para los períodos en estudio de la comunidad.
- d) *Presentar resultados:*
 - a. *Definir objetivos de la comunidad:* Los administradores deben definir los niveles de moderación que quieren para la comunidad, en términos de qué fracción del total de interacciones entre los miembros de la comunidad debiera ser moderada.
 - b. *Evaluar los niveles de moderación:* Se calculan los niveles de moderación de la comunidad, para todo el periodo en el cual se encuentra activa. Se generan semáforos para mostrar si el nivel de moderación difiere de los niveles definidos por los administradores.

Capítulo 5 – Implementación de la Solución

Con el objeto de automatizar el proceso de evaluación, de los sistemas de software utilizados por las comunidades virtuales en estudio, es que se ha desarrollado un sistema de análisis y reportes de comunidades virtuales. Este sistema básicamente toma la información de las base de datos de los software utilizados por una determinada comunidad virtual, transfiere la información a un modelo de datos común, calcula un conjunto de métricas preestablecidas, y las despliega en reportes a través un sitio Web. A continuación se detalla cómo está implementado.

5.1 Arquitectura General

En la Figura 18 se muestra la arquitectura general del sistema implementado. Como se puede observar, existen varios repositorios de datos, así como varios programas que contribuyen al sistema en general.

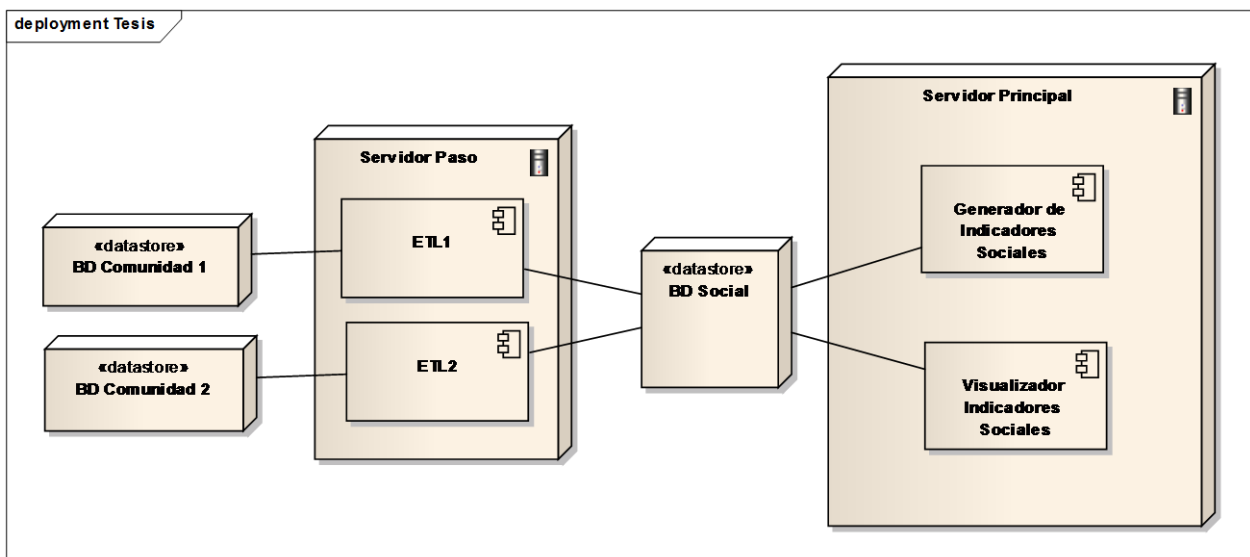


Figura 18. Arquitectura del sistema de procesamiento y análisis.

Entre los repositorios tenemos:

- (a) Las Bases de Datos de los sistemas de software de las comunidades virtuales en estudio por lo general se encuentran en mysql o postgres. Cada sistema posee su modelo de datos propios. En algunos casos, también se cuenta con el log de los accesos al sistema web.

- (b) La Base de Datos Social, modelado según la especificación del Capítulo 2.4.3. Aquí se graban todos los datos en una misma estructura, la cual permite utilizar los mismos programas para las diferentes comunidades. También aquí se almacenan los indicadores después de haber sido calculados.

Dentro de los programas que hay, están:

- (a) ETL de conversión desde las bases de datos de cada sistema, al repositorio central. Para cada tipo de software es necesario crear uno. Al respecto, para poder hacer la conversión, lo principal es poder determinar cómo interactúan los miembros que utilizan dichos sistemas de software, para determinar cómo efectuar un ETL. Por ejemplo, tal como se puede ver en la Tabla 2, en un foro de discusión un Contexto es un tema de discusión, mientras que una Interacción es un Post.

Tabla 2. Contextos e Interacciones para distintos sistemas de Software.

Tipo de Software	Contexto	Interacción
Foro de Discusión	Tema de Discusión	Post
Blog de Noticias	Noticia	Comentarios
Twitter	Tweet	Respuestas
Foro: Mensajes Privados	Mensaje Privado	Respuestas

- (b) Generador de Indicadores Sociales. Programa que implementa las técnicas de evaluación de los indicadores sociales.
- (c) Programa para desplegar los indicadores sociales en forma de reportes. Algunas consideraciones sobre los reportes Web:
- El reporte que se genere debe ser simple de comprender. Algunos de los administradores ya cuenta con experiencia en la visualización de reportes Web, por lo cual si el reporte es similar facilitaría su lectura.
 - La generación y visualización del reporte no debe afectar el normal uso de la herramienta de foro de discusión al interior de la comunidad. Por tanto, si el reporte no es absolutamente en línea, no es un problema.

5.2 Tecnologías Usadas en la Implementación

Para la implementación del generador de métricas se utilizó:

- Java versión 1.6.
- Jung 2.2.0: como librería para calcular métricas de redes sociales.

- Commons Math 3.0: librería para efectuar las regresiones sobre el conjunto de datos, y determinar las distribuciones.
- JGibbLDA: Implementación Java de Latent Dirichlet Allocation (LDA).

Para la implementación del reporte vía Web se utilizó:

- PHP, bajo un servidor Apache.
- jquery y angularJS, para la interfaz basada en Ajax.

5.3 Despliegue de los Indicadores

La información que es procesada para cada comunidad es desplegada posteriormente en un *dashboard*, tal como la siguiente figura:

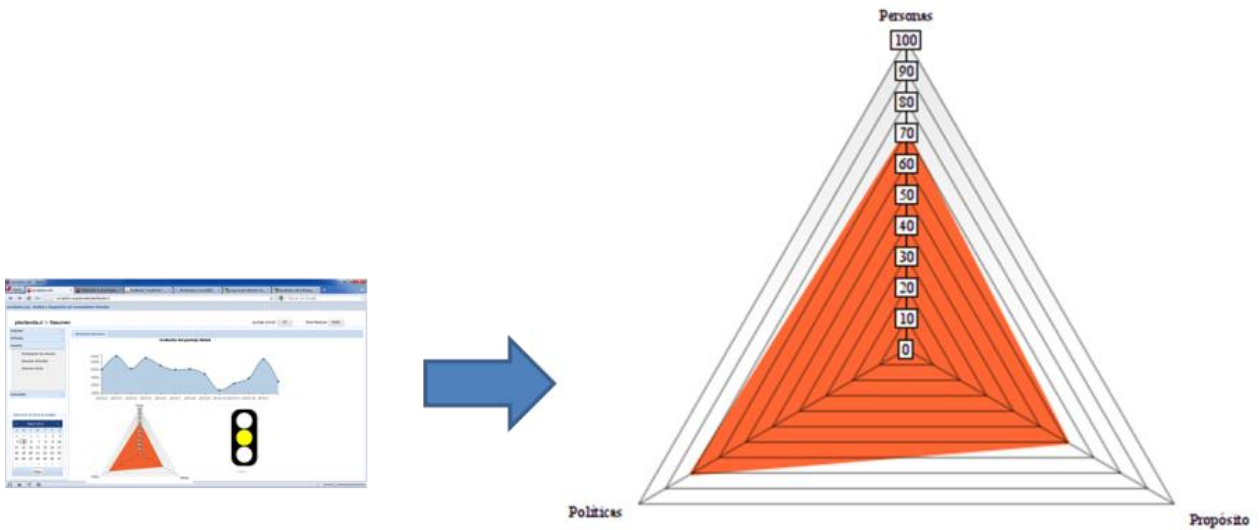


Figura 19. Pantalla con despliegue de indicadores.

Capítulo 6 – Estudio de Comunidad I

Ésta corresponde a una Comunidad Virtual de Práctica formada por un grupo de personas que se han reunido con el objetivo de compartir conocimiento y experiencias en la construcción de efectos de música, amplificadores y equipos de audio. Ellos se caracterizan por tener un enfoque hacia “hazlo tú mismo” (DIY – Do It Yourself). En un comienzo los miembros de la comunidad se encontraban enfocados principalmente en la construcción y utilización de plexies (el cual corresponde a un amplificador clásico con el cual es posible obtener un sonido diferente al que se obtiene con amplificadores tradicionales). Sin embargo, se formaron nuevos temas con el transcurso de los años, tales como la *lutería* (corresponde a la confección de instrumentos musicales de cuerda, como por ejemplo, una guitarra), audio profesional, compra/venta de partes, etc.

La comunidad virtual cuenta con aproximadamente 14 años de existencia, en los cuales han participado más de 2.500 miembros. La interacción entre los miembros de la comunidad es realizada a través de un sistema de foro de discusión. En la actualidad se encuentra en una etapa en la cual la participación de sus miembros ha ido decayendo considerablemente en los últimos años. Al comienzo de la existencia de la comunidad la tarea de administración era una tarea más bien sencilla y era efectuada por un solo miembro. En el período de mayor actividad de la comunidad virtual dicha tarea alcanzó a ser realizada por varios administradores.

6.1. Análisis de la Comunidad

Se dispone de los datos de comunidad, desde sus comienzos el año 2002, hasta mediados del año 2015. Los miembros de esta comunidad disponen básicamente de dos formas de interactuar con otros miembros (a) a través de la publicación de mensajes al interior de uno de los foros de discusión de la comunidad, y (b) a través del envío de mensajes privados hacia otros miembros de la comunidad. Tal como es posible visualizar en la Figura 20, la mayor cantidad de interacciones se da entre los años 2007 y 2008. El estudio de esta comunidad se encuentra enfocado en los mensajes que se publican en los diferentes foros de la comunidad, los cuales corresponden a casi la totalidad de las interacciones entre los miembros de la misma.

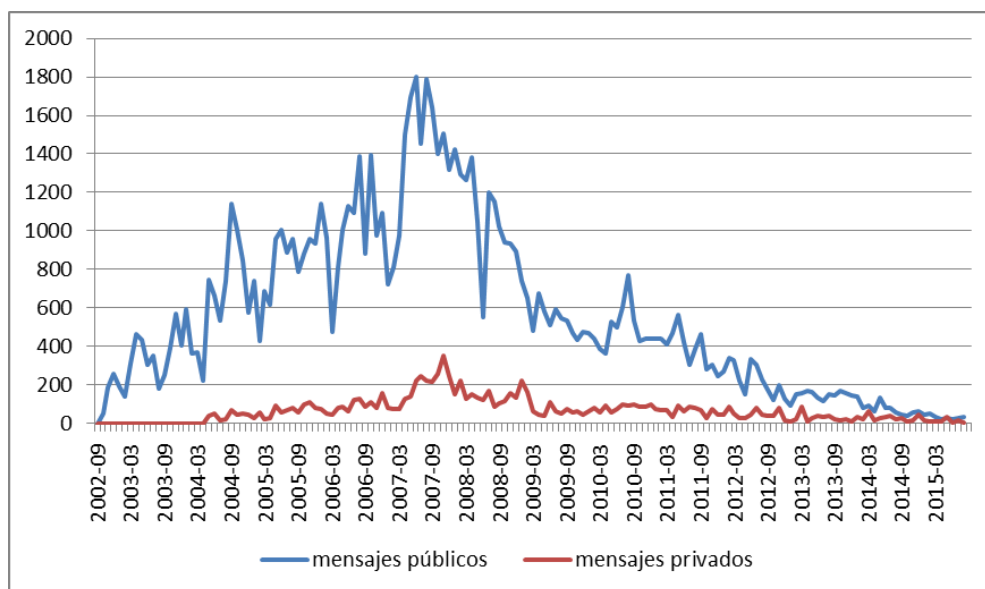


Figura 20. Actividad de la Comunidad I.

En la Tabla 3 es posible visualizar cómo se distribuyen las participaciones de los miembros de la comunidad en los diferentes foros de discusión, ya sea (a) creando un nuevo hilo de discusión (llamados *topics*), o (b) escribiendo un mensaje en un hilo de discusión existente (llamados *posts*).

Tabla 3. Participación de los miembros de la Comunidad I.

Foro	# posts	# topics	%
<i>Amplificadores</i>	22.918	2.705	26,3%
<i>Effectos</i>	32.062	3.470	36,8%
<i>Luthería</i>	10.083	1.399	11,6%
<i>General</i>	19.886	2.707	22,8%
<i>Audio Pro</i>	1.850	201	2,1%
<i>Sintetizadores</i>	327	35	0,4%
<i>Hardcore DIY</i>	117	12	0,1%
Total	87.243	10.529	100%

La mayoría de los foros fueron creados en los inicios de la existencia de la comunidad virtual, con excepción de los foros de “*Audio Pro*”, “*Sintetizadores*” y “*Hardcore DIY*”, los cuales fueron creados en los últimos años, contando este último foro con acceso sólo para ciertos miembros de la comunidad.

Otro aspecto importante que es posible visualizar en esta comunidad virtual, es que la participación en el foro “General” (creado para discusiones generales, no necesariamente relacionadas con el objetivo de la comunidad virtual) alcanza el 22,8% del total de interacciones, mientras que la participación en los otros foros (relacionados con temas específicos de la comunidad virtual) corresponde al 77,2%.

En la Figura 21 se despliega la cantidad de miembros que participan activamente en la comunidad a través de los diferentes años de existencia de la comunidad virtual. Es posible observar que la mayor cantidad de miembros activos se encuentran entre los años 2006 y 2009.

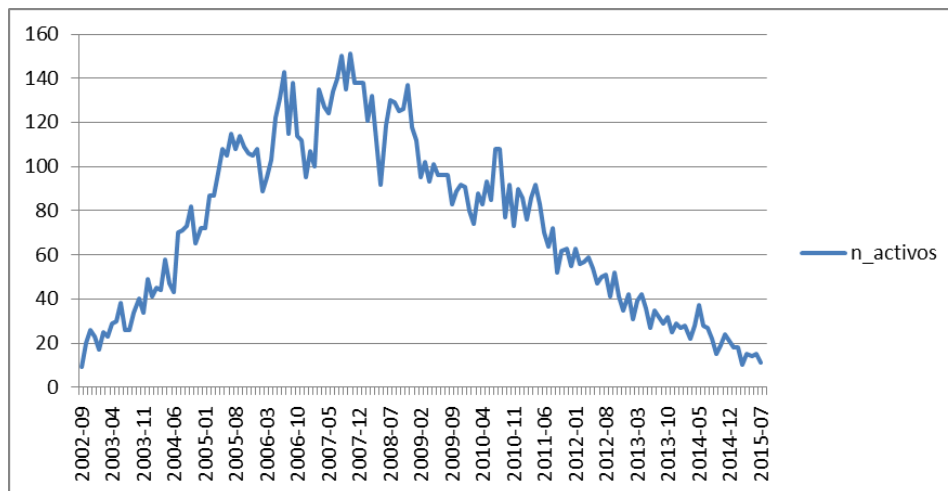


Figura 21. Evolución de los niveles de participación de la Comunidad I, basados en el total de interacciones.

6.2 Proceso de Extracción de Datos

En la Figura 22 es posible ver el diagrama de clases simplificado del sistema de software utilizado por los miembros de la comunidad virtual. A continuación se describen brevemente estas clases:

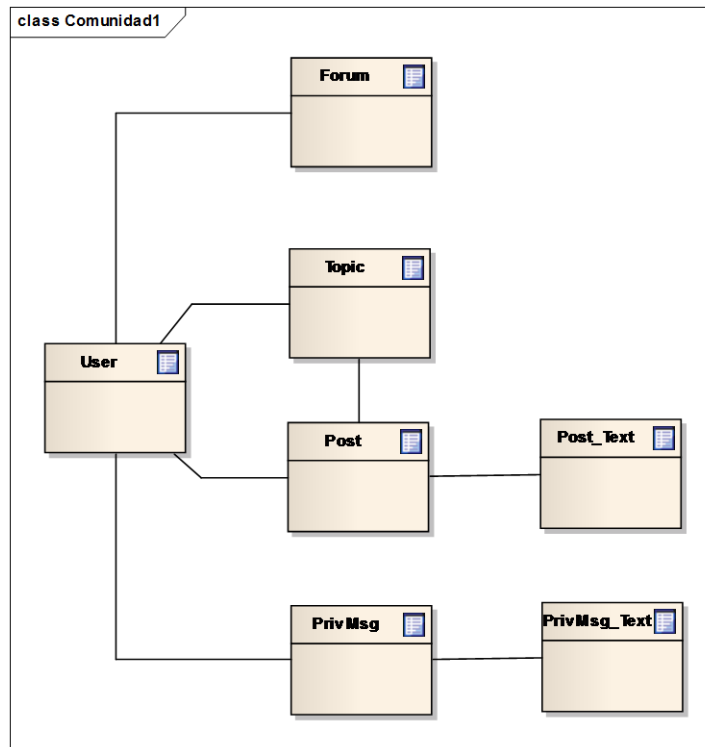


Figura 22. Diagrama de clases simplificado del Foro de Discusión utilizado por la Comunidad I.

Forum: Corresponde a los foros de discusión que son creados por un administrador. Tal como fue mencionado anteriormente, existe un Foro denominado “*General*” y otros específicos a cada tema que es tratado al interior de la comunidad virtual.

Topic: Corresponde a los hilos de discusión que son creado en cada foro por los miembros de la comunidad. Un hilo de discusión debe ser creado al interior de un foro de discusión. En caso de que el hilo de discusión fue creado en el foro de discusión equivocado, un administrador puede moverlo al foro de discusión correcto. En caso de que el hilo de discusión trate de un tema completamente fuera de la comunidad, o que no cumpla con el reglamento de la comunidad, puede ser bloqueado, con lo cual se impide que otros miembros de la comunidad sigan publicando mensajes dentro de dicho hilo de discusión.

Post: Corresponde a los mensajes de cada hilo de discusión. Son escritos por los miembros de la comunidad, para lo cual debe seleccionar un hilo de discusión previamente creado. Pueden ser editados por los administradores de la comunidad.

Post_text: Corresponde al texto del mensaje de un determinado post.

User: Corresponde a las personas que son miembros de la comunidad virtual. Pueden crear hilos de discusión al interior de un foro (*Topic*) o publicar mensajes al interior de un hilo de discusión existente (*Post*). Un administrador también es un usuario (*User*), lo cual es indicado en uno de

sus atributos. Un usuario puede ser bloqueado por un administrador si su conducta es inadecuada al interior de la comunidad virtual.

Privmsgs: Corresponden a los mensajes que son enviados desde un miembro de la comunidad virtual, hacia otro. Son utilizados comúnmente para solicitar ayuda al administrador.

Privmsgs_Text: Corresponden al texto de un determinado mensaje privado.

En la siguiente tabla se muestra la correspondencia entre los conceptos de esta comunidad, y el modelo social propuesto. En base a esta información se crea el proceso que extrae los datos de esta comunidad virtual hacia el modelo de datos común, en base al cual se aplican las diferentes técnicas de evaluación.

Tabla 4. Correspondencia entre conceptos de la Comunidad I y el modelo social propuesto.

Concepto Social	Comunidad Virtual I
Personas	Usuarios
Grupos	No Aplica
Roles	Administrador o Usuario, de acuerdo al atributo <code>is_admin</code> en la tabla Usuarios
Interacción	Post
Contexto	Foros + Tópicos
Contenido	Post_Text
Tags	No Aplica
Permisos	Usuarios / Tipo
Moderación Persona	Atributo <code>is_bloqued</code> en la tabla Usuarios, el cual indica si un usuario fue bloqueado por un administrador
Moderación Interacción	Atributo <code>is_edited</code> en la tabla Posts, el cual indica que el mensaje escrito por un Usuario, fue editado por un Administrador.
Moderación Contexto	Atributo <code>is_bloqued</code> en la tabla topics, el cual indica que todo un hilo de discusión fue bloqueado por un administrador. Atributo <code>is_moved</code> en la tabla topics, el cual indica que todo un hilo de discusión fue movido a otro foro por un administrador.

6.3 Evaluación de la Estrategia Propuesta – Personas

A continuación, para cada una de las comunidades en estudio, se muestra la aplicación de la estrategia propuesta para el aspecto social de *personas*.

6.3.1 Preparación de Datos

El primer paso es la generación de las redes sociales en base a los datos de las interacciones de los miembros de la comunidad. Tal como anteriormente, se generan cuatro tipos de redes sociales para cada periodo en estudio: (a) *reply_all*, (b) *reply_prev*, (c) *reply_all30*, y (d) *creator*. El estudio se realiza a nivel mensual, por solicitud de los administradores de la comunidad, por lo que son 168 periodos (doce meses para cada año comprendido entre el 2002 y 2015), con lo cual se generan 672 redes sociales.

Es posible observar (en la Figura 23) que las redes sociales que se forman poseen diferentes niveles de densidades a través de los años de existencia de la comunidad virtual, siendo las redes *reply_all30* y *reply_all* las más densas.

The figure consists of four subplots, each showing a line graph of network density over time. The y-axis for all graphs ranges from 0 to 1.0 in increments of 0.1. The x-axis represents time, with 168 data points corresponding to 14 years of monthly data. All graphs show a general downward trend from a high initial density (around 0.9) to a lower, more stable density (around 0.1-0.2) over the period. Graph (a) *reply_all* shows a steady decline with some fluctuations. Graph (b) *reply_prev* shows a similar trend but with more pronounced fluctuations. Graph (c) *reply_all30* shows the highest initial density and the most significant fluctuations, with several peaks reaching above 0.5. Graph (d) *creator* shows the lowest initial density and the most stable, lowest overall density throughout the period.

Figura 23. Evolución de las densidades para las redes sociales de la Comunidad I: (a) *reply_all*, (b) *reply_prev*, (c) *reply_all30*, (d) *creator*.

56

Si bien es posible visualizar las redes sociales que se generan (Figura 24), el objetivo es obtener indicadores que sean más fáciles de interpretar por los administradores de la comunidad virtual en estudio por lo cual no son utilizadas mayormente dichas visualizaciones.

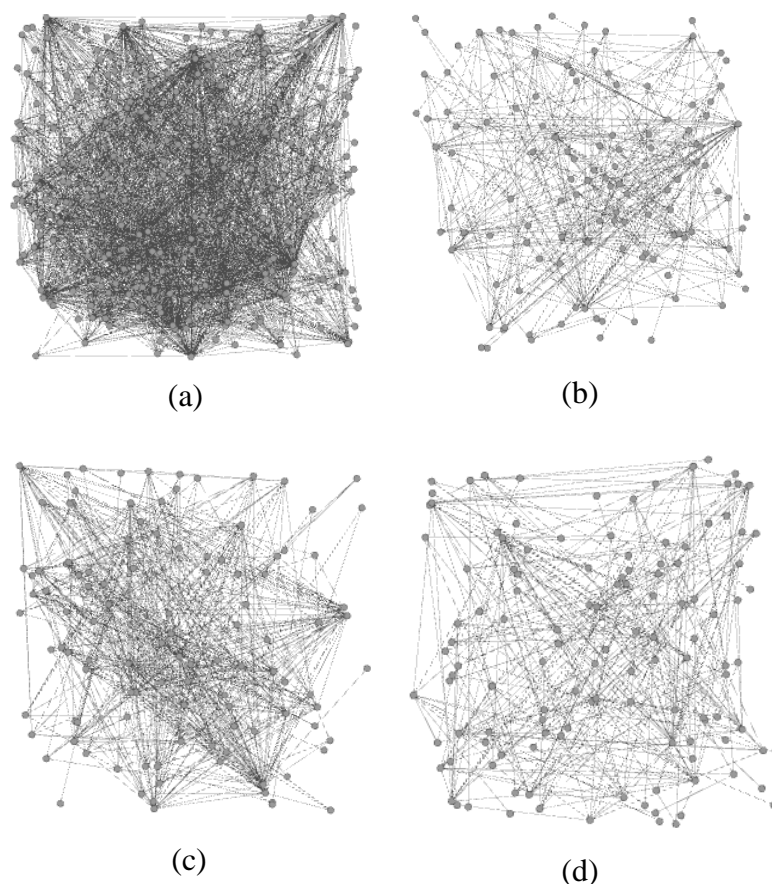


Figura 24. Redes sociales para enero del año 2009 (2009-01) de la Comunidad I: (a) *reply_all*, (b) *reply_prev*, (c) *reply_all30*, (d) *creator*.

6.3.2 Aplicación de las Técnicas de Evaluación

A continuación se explica cómo se aplican las técnicas de evaluación en el cálculo de los diferentes niveles de participación:

(a) Cálculo de Miembros Activos y Núcleo

Para cada red social generada (672 en total), se calcula la centralidad de sus nodos en base a la técnicas de degree, betweenness centrality, closeness centrality, hits (authority y hub), y pagerank. Para cada resultado obtenido, se calcula la cantidad de miembros que pertenecen al núcleo y los que pertenecen a los miembros activos. Específicamente, para calcular la cantidad de miembros pertenecientes al núcleo, se dibuja la distribución del índice de centralidad de los miembros de la

comunidad, se calcula la regresión que mejor se ajuste a los datos obtenidos (en la mayoría de los casos se aproxima a una distribución de potencia, y en los otros a una distribución exponencial), y se calcula el “codo” de la curva que representa a la regresión. Para calcular los miembros activos, se consideran a todos los usuarios que tengan un índice de centralidad mayor que 0. Un extracto de los resultados obtenidos es posible visualizarlos en la siguiente tabla:

Tabla 5. Resumen de los datos obtenidos para la Comunidad I en el cálculo de los miembros pertenecientes al núcleo.

	degree				Hits / Authority			
	reply_all30	reply_all	creator	reply_prev	reply_all30	reply_all	creator	reply_prev
2009-01	19	26	11	13	9	9	8	9
2009-02	21	23	10	11	7	7	7	8
2009-03	23	26	12	14	9	8	8	9
2009-04	20	22	10	14	9	8	8	9
2009-05	20	24	9	12	8	8	7	8
2009-06	21	22	10	13	8	7	8	8
2009-07	20	23	10	13	8	8	8	8
2009-08	18	21	10	12	8	8	7	9
2009-09	17	18	9	12	7	7	7	8
2009-10	17	19	8	10	7	7	6	8
2009-11	18	19	8	11	8	8	6	8
2009-12	17	19	8	12	8	7	7	7
2010-01	17	18	8	11	7	7	7	7
2010-02	18	19	8	11	7	7	6	8
2010-03	15	16	7	9	6	6	5	7
2010-04	16	17	9	12	8	8	6	8
2010-05	17	19	9	12	8	8	7	8
2010-06	21	24	11	14	8	8	8	9
2010-07	25	26	12	15	10	10	9	10
2010-08	18	21	9	12	8	7	7	8
2010-09	20	22	8	11	8	8	7	8
2010-10	16	18	8	11	8	8	6	8
2010-11	17	18	9	11	7	7	6	8
2010-12	18	20	9	11	7	7	6	8

(b) Cálculo de periferia

Para calcular la cantidad de miembros que pertenece a la periferia, es necesario identificar a los miembros que no publican en un periodo dado, pero que aún son parte de la comunidad. Para esta comunidad se calcula en base a la información de la fecha de registro en la comunidad y la fecha de última actividad. Con dicha información es posible obtener el total de miembros que pertenece a la comunidad en un periodo dado, incluyendo a aquellos miembros que participan activamente y a aquellos que no (ver Figura 25).

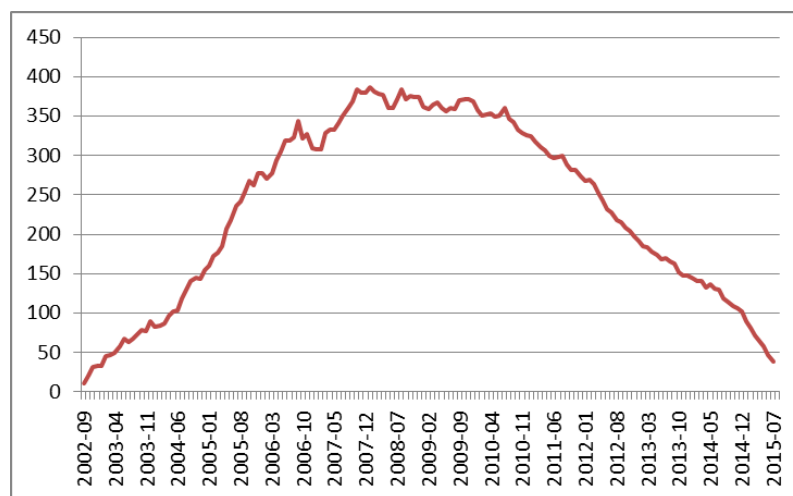


Figura 25. Miembros que pertenecen a la periferia en la Comunidad I.

6.3.3 Evaluación de los Resultados

A continuación se explica cómo se evalúan los resultados obtenidos, de acuerdo a lo especificado en la estrategia de evaluación.

(a) Cálculo de precisión y recall.

Para poder estimar la precisión de las técnicas utilizadas sobre las diferentes redes sociales, se les solicitó a los administradores de la comunidad que identifiquen a los miembros que ellos consideraban que son los más importantes en un periodo dado. Dicha información es contrastada posteriormente con los datos obtenidos para cada técnica de evaluación en la estimación de aquellos miembros que pertenecen al núcleo de la comunidad virtual. Se cuenta con esta información para dos periodos:

- Año 2014, la cual fue entregada a principios del año 2015 (con motivo de la escritura del presente documento)
- Año 2009-2010, la cual fue entregada a fines del año 2010 (fecha en la cual se hizo el primer estudio de esta comunidad virtual)

Con la información entregada es posible calcular el índice F para cada técnica aplicada. En la Figura 26 es posible visualizar el resultado obtenido para la técnica *degree* en el periodo 2009-2010. Es posible observar que se obtienen los mejores resultados al utilizar las redes sociales *reply_all* y *reply_all30*.

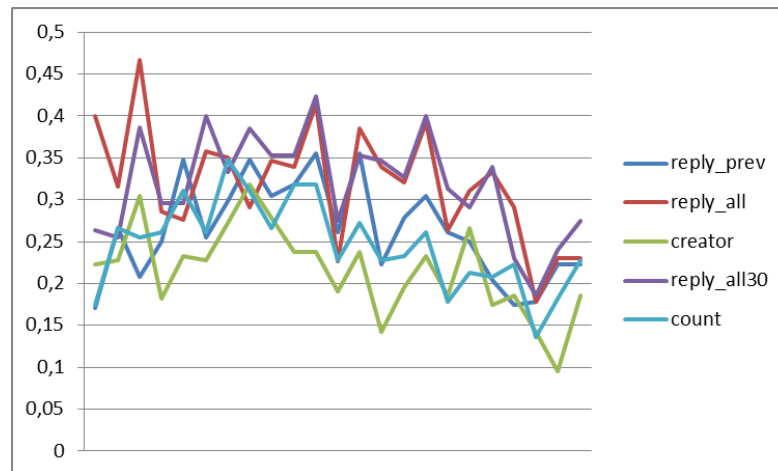


Figura 26. Evolución de índice F para el periodo 2009-2010 de la Comunidad I, utilizando *degree* como medida de centralidad.

En la Figura 27 es posible observar la comparación de los resultados obtenidos para la red social *reply_all* utilizando las diferentes técnicas. Es posible concluir que el mejor resultado se obtiene utilizando la técnica *degree*.

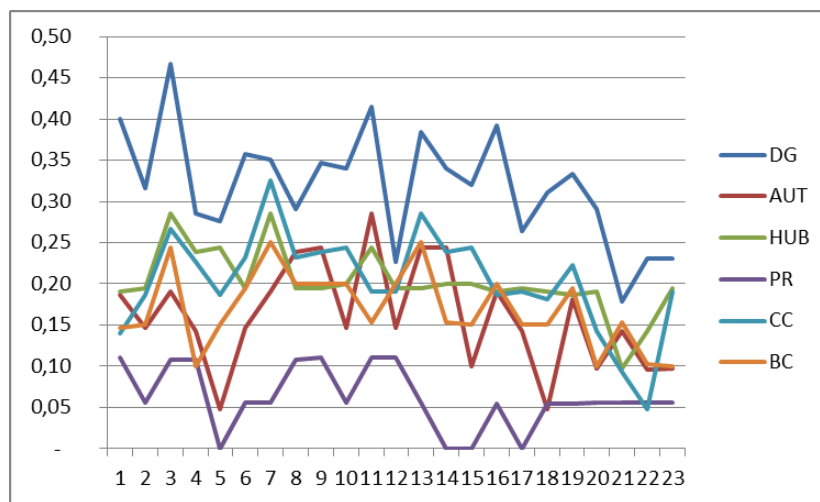


Figura 27. Comparación del índice F de la Comunidad I, para las 6 técnicas aplicadas, utilizando la red social *reply_all*.

Con el mejor resultado obtenido, *degree* con la red social *reply_all*, es posible calcular los niveles de participación para el resto todos los periodos (ver Figura 28).

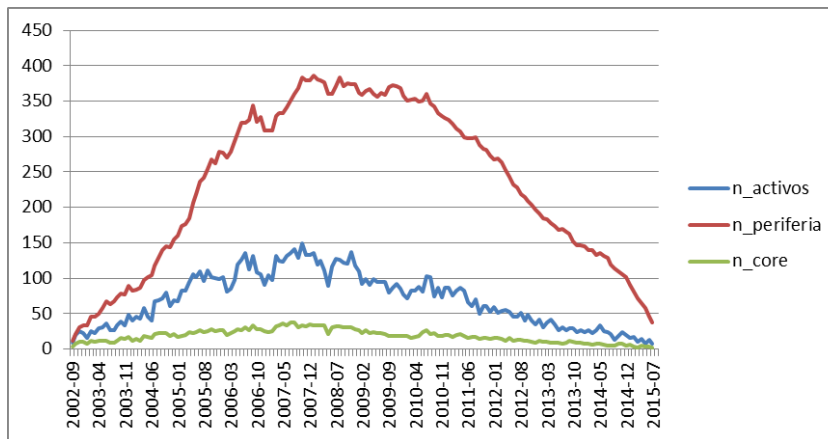


Figura 28. Niveles de participación para la Comunidad I, tomando como métrica la técnica *degree con reply_all*.

(b) Comparación con conteo simple

Uno de los aspectos que es necesario evaluar, es si el haber aplicado técnicas de SNA genera mayor precisión en la estimación de los niveles de participación, que el hecho de haber utilizado solamente la cantidad de publicaciones por usuario como técnica de evaluación (a lo cual denominaremos “conteo simple”, y que corresponde a la métrica utilizada por los administradores de la comunidad).

Para lo anterior se calcula los niveles de participación de la siguiente forma: (a) *miembros activos*: aquellos miembros que publican un mensaje en el periodo en estudio, (b) *núcleo*: calculado en base a la distribución de participaciones, y (c) *periferia*: calculada de la misma forma que para las técnicas anteriores, es decir, en base a las fecha de registro en la comunidad y de último acceso. De esta forma, se obtiene los niveles de participación que se muestran en la Figura 29.

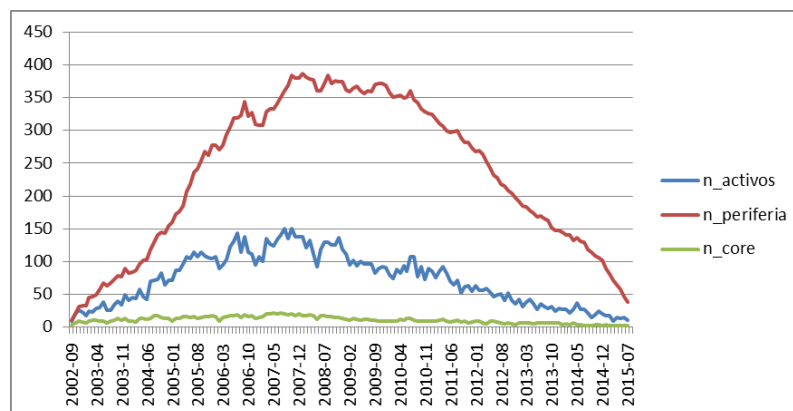


Figura 29. Evolución de los niveles de participación de la Comunidad I, basados en el total de interacciones.

Es posible visualizar que el tamaño del núcleo es menor que el obtenido a partir de las técnicas de SNA. En la Figura 30 es posible visualizar el tamaño del núcleo y de los miembros activos, en forma porcentual.

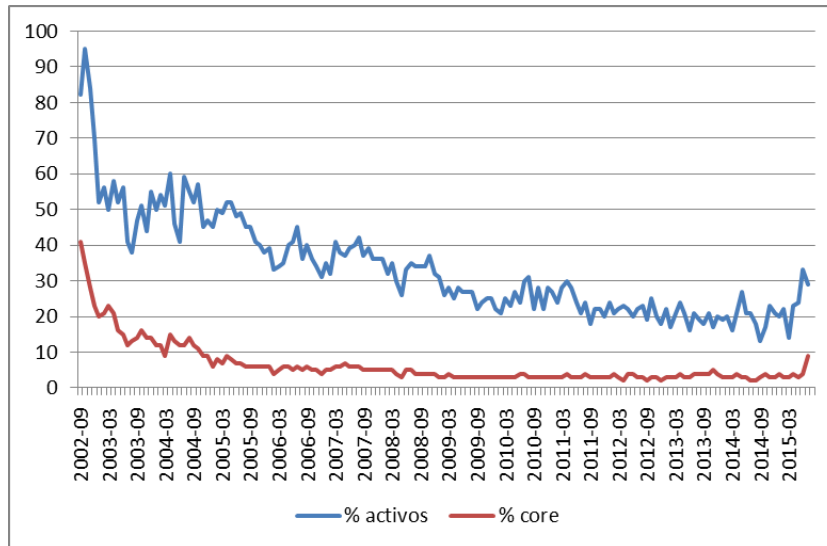


Figura 30. Evolución de los niveles de participación en forma porcentual de la Comunidad I, basados en el total de interacciones.

Si se compara los resultados obtenidos de esta forma, con el mejor resultado obtenido en base a las técnicas de SNA (*degree* sobre la red *reply_all*), es posible visualizar que se obtienen mejores resultados utilizando la técnica de SNA (Figura 31).

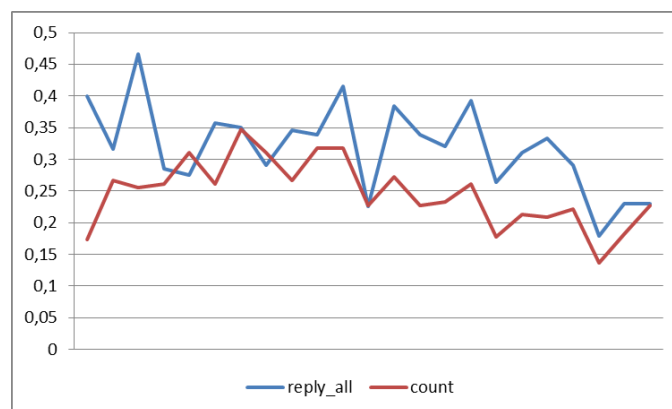
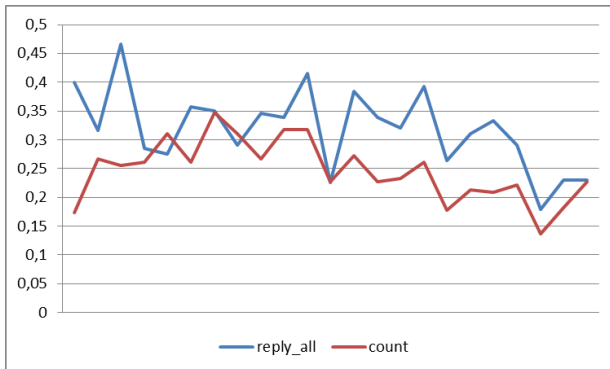
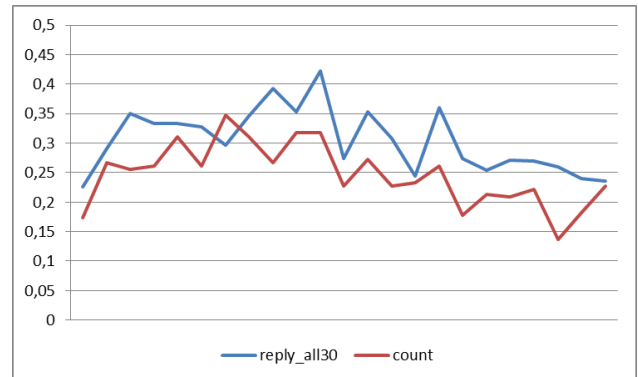


Figura 31. Evolución de índice F para el periodo 2009-2010 de la Comunidad I, utilizando *degree* como medida de centralidad.

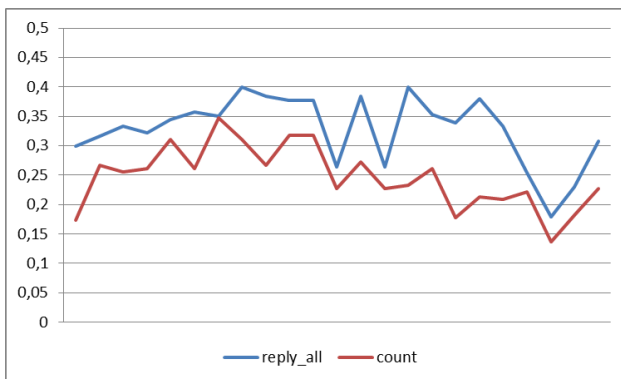
Si se compara el resultado obtenido por conteo simple, con cada una de las técnicas utilizadas (ver Figura 32), es posible visualizar que, en general, el conteo simple sólo es mejor respecto de las técnicas de *closeness* y *betweness centrality*, y es peor respecto de las demás técnicas.



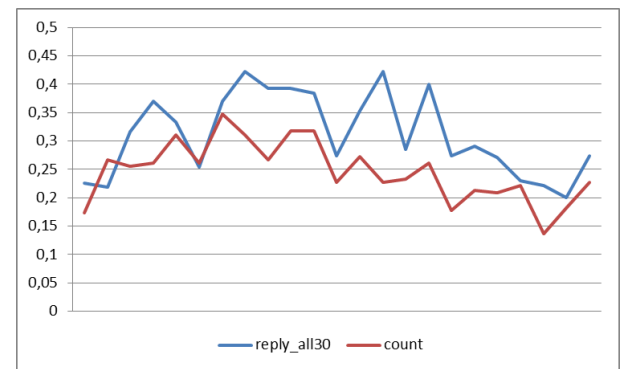
(a) *degree*



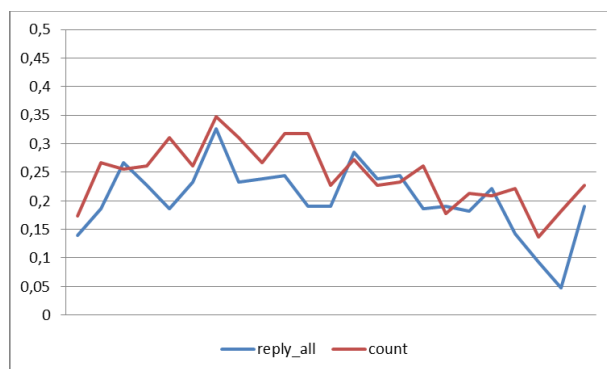
(b) *hits - authority*



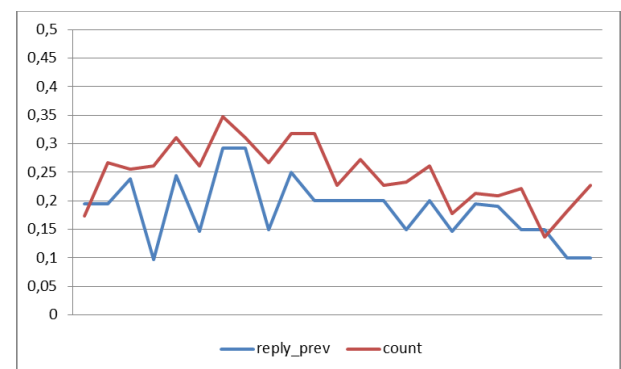
(c) *hits - hub*



(d) *pagerank*



(e) *closeness centrality*



(f) *betweness centrality*

Figura 32. Comparación entre las diferentes técnicas SNA y conteo simple en la Comunidad I.

6.3.4 Presentación de los Resultados

Los resultados son presentados de dos formas: (a) en forma temporal, para poder comprender cómo evoluciona a través del tiempo los niveles de participación, y (b) como un semáforo, para poder indicar en un periodo determinado el nivel de salud de la comunidad, respecto de la evaluación de este aspecto social. En base a la técnica *degree* con la red social *reply_all*, se calcula la evolución temporal de los niveles de participación, contando los miembros que pertenecen a cada nivel de participación (Figura 33), y calculándolo en forma porcentual respecto del total de miembros de la comunidad en un instante de tiempo determinado (Figura 34),

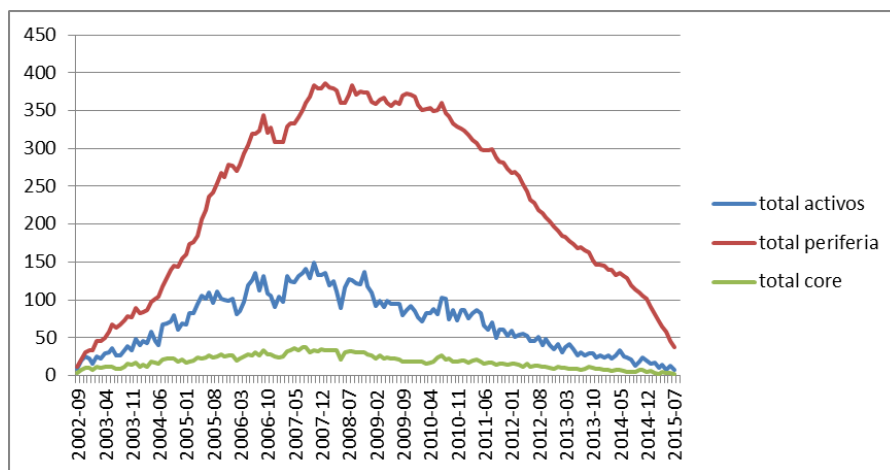


Figura 33. Evolución temporal de los niveles de participación para la Comunidad I.

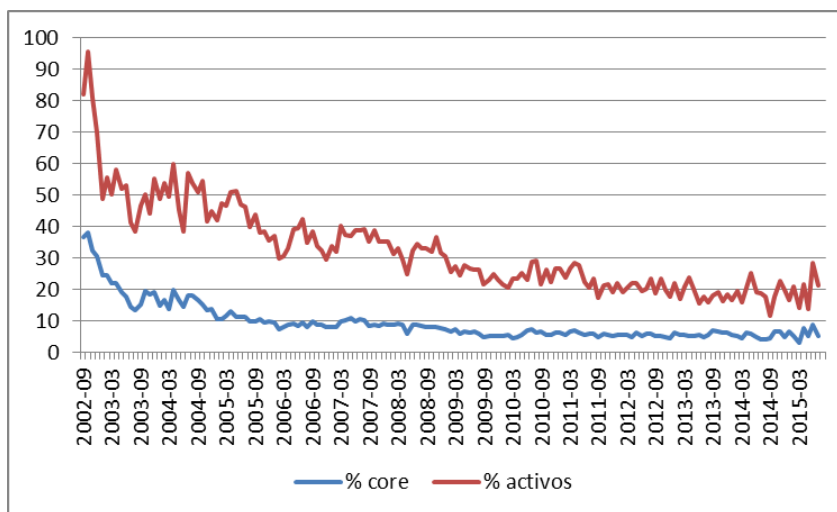


Figura 34. Evolución de los niveles de participación para la Comunidad I, en forma porcentual.

Según la definición de los administradores de la comunidad, se utilizan los rangos indicados en la Tabla 6 para determinar si los niveles de participación son adecuados o no.

Tabla 6. Rangos para determinar si niveles de participación son adecuados para la Comunidad I.

Color	Núcleo	Miembros Activos
Rojo	< 8%	< 20%
Amarillo	~ 8-10 %	20-25 %
Verde	> 10 %	> 25 %

Con la definición anterior, es posible visualizar (Figura 35) un resumen de los niveles de participación.

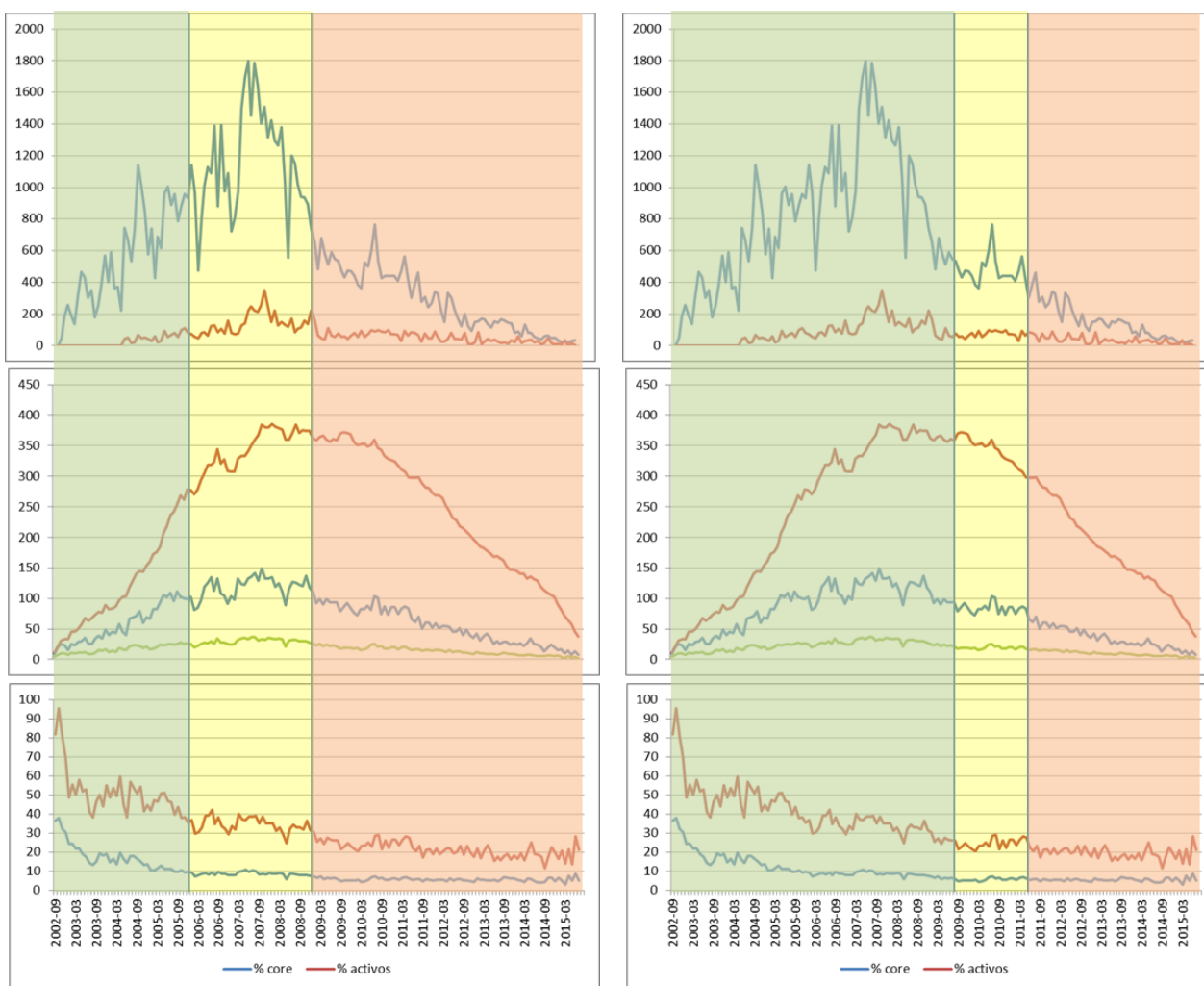


Figura 35. Resumen de los niveles de participación de la comunidad I: (a) Resumen miembros pertenecientes al núcleo de la comunidad I, (b) Resumen miembros activos de dicha comunidad.

Es posible observar que durante el periodo de inicial de la comunidad, el nivel de participación del núcleo y de los miembros activos se encontraba “saludable”. Sin embargo, en el periodo de mayor participación de los miembros de la comunidad, el núcleo de la comunidad fue disminuyendo. En cambio, el nivel de participación de los miembros activos se fue reduciendo en conjunto con la participación completa de los miembros de la comunidad virtual.

6.3.5 Resultado de la Encuesta

La encuesta para medir el indicador de *identidad* fue respondida por 9 miembros de la comunidad, cuyos resultados se encuentran en la Tabla 7. La encuesta fue realizada en el período 2015-07, siendo respondida por el 64% de los miembros activos en dicho período. Dado que no

hay información de la encuesta para los períodos pasados, es que los resultados de esta encuesta se utilizan sólo como una referencia.

Tabla 7. Resultado de la Encuesta de Identidad para la Comunidad I.

Pregunta	Muy en desacuerdo			Muy de Acuerdo		PROMEDIO
	1	2	3	4	5	
1			2		7	4,6
2				1	8	4,9
3		1	4	2	2	3,6
4		3	4	1	1	3,0
						4,0

Tal como se puede ver, en general la evaluación es buena. La nota más baja se obtiene con la pregunta “¿Es importante para ti ser un miembro de la comunidad X?”.

6.4. Evaluación de la Estrategia Propuesta - Propósito

A continuación, para cada una de las comunidades en estudio, se muestra la aplicación de la estrategia propuesta para el aspecto social de *propósito*.

6.4.1 Preparación de los Datos

Para poder aplicar las técnicas de evaluación es necesario extraer los textos de las diferentes participaciones de los miembros de la comunidad. Dichos textos se agrupan por contextos de interacción, los cuales en este caso, corresponden a diferentes hilos de discusión (previamente se realizó el análisis sin agrupar las participaciones, pero no se generan mejores resultados, debido a que muchas de las participaciones están referidas a un contexto de conversación específico, y por lo tanto, omiten palabras claves que ya se encuentran mencionadas en otros posts dentro del mismo hilo de discusión).

Después del proceso de limpieza de los textos, se generan los archivos de acuerdo al formato que requieren las técnicas de evaluación: (a) un archivo de texto por período de estudio, (b) en cada archivo de texto, cada línea representa un documento, (c) la primera línea indica el total de documentos, y (d) cada línea tiene las palabras de un documento específico.

Para este análisis cada periodo corresponde a un año completo, y cada documento corresponde a un hilo de discusión. En la siguiente Tabla se muestra un resumen de los archivos generados para su posterior análisis.

Tabla 8. Resumen de los documentos para analizar de la Comunidad I.

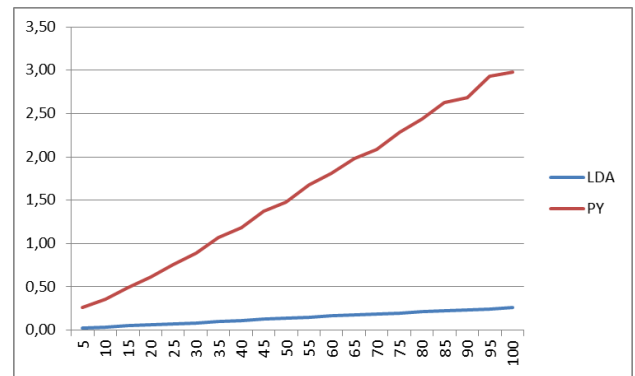
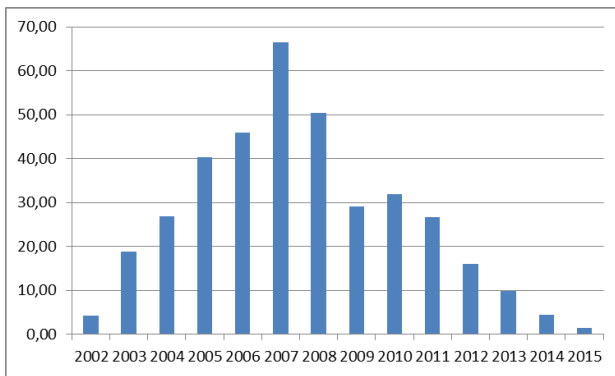
Año	total docs	total words	promedio
2002	84	6.384	76
2003	471	18.184	38,6072187
2004	916	26.696	29,1441048
2005	1.487	32.513	21,8648285
2006	1.690	36.601	21,6573964
2007	1.883	49.974	26,5395645
2008	1.415	41.270	29,1660777
2009	969	27.751	28,6388029
2010	876	29.285	33,4303653
2011	646	25.317	39,1904025
2012	412	18.368	44,5825243
2013	277	13.137	47,4259928
2014	197	7.321	37,1624365
2015	59	3.089	52,3559322

6.4.2 Aplicación de las Técnicas de Evaluación

Con los archivos generados previamente, se aplican las técnicas de evaluación con diferentes configuraciones, de tal forma de poder determinar cuál técnica genera mejores resultados, y cuál es la cantidad óptima de tópicos por cada periodo en estudio.

Dado que las técnicas de LDA y PYTM requieren como parámetro la cantidad de tópicos, las distintas configuraciones irán cambiando el total de tópicos que se debe calcular. Para ello, se calcula dichas técnicas desde 0 hasta 100 tópicos, pero con incrementos de 5 (debido a los tiempos que tarda la técnica en calcular). Adicionalmente, como se tiene sospecha de que la cantidad óptima de tópicos es menor a 25, se calculará también desde 1 a 25 tópicos, con incrementos de 1, de tal forma de que, al momento de encontrar la cantidad óptima de tópicos, está ya se encuentre calculada sobre el conjunto de documentos. Por lo tanto, dado que son 14 periodos en estudio (desde el año 2002 al 2015), y 34 experimentos (del 5 al 100 en incrementos de 5, y del 1 al 25 con incrementos de 1), son en total 476 ejecuciones de cada técnica.

A diferencia de las técnicas del capítulo anterior, éstas tardan significativamente más tiempo en ser procesadas, es especial la técnica de PYTM. A continuación, en la Figura 36, se despliegan los tiempos de procesamiento de dicha técnica. Sólo para esta técnica, son más de 372 horas de procesamiento. En la misma figura es posible visualizar que la técnica PYTM tarda mucho más tiempo que LDA, a medida que se calcula una mayor cantidad de tópicos.



(a)

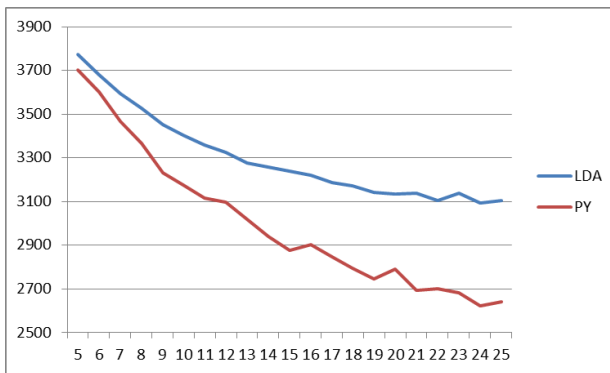
(b)

Figura 36. Resumen de los tiempos de procesamiento de LDA y PYTM en la Comunidad I: (a) Tiempos de procesamiento de PYTM, medido en horas, entre los años 2002 y 2015 (b) Comparación de tiempos de procesamiento entre LDA y PYTM, para el año 2010, con cantidad de tópicos entre 5 y 100, medido en horas.

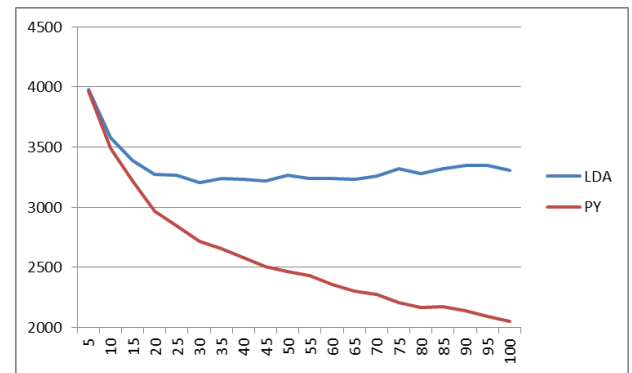
6.4.3 Evaluación de los Resultados

(a) Selección de técnica de evaluación

Tal como es mencionado en los capítulos anteriores, para evaluar los resultados se calcula la métrica *perplexity* para cada aplicación de la técnica. Como es posible ver en la Figura 37, la técnica de PYTM entrega mejores valores para los diferentes periodos en estudio.



(a)



(b)

Figura 37. Comparación de técnicas LDA y PYTM en la Comunidad I: (a) Año 2004, entre 5 y 25 tópicos, comparación entre LDA y PYTM, (b) Año 2010, entre 5 y 100 tópicos, comparación entre LDA y PYTM.

(b) Selección de cantidad de tópicos

Para la técnica seleccionada anteriormente, se analiza para cada periodo la cantidad óptima de tópicos. Para ello, se normalizan los valores de la métrica, y sobre la curva de la métrica normalizada se calcula la cantidad optima sobre la regresión sobre los valores obtenidos (en todos los periodos se aproxima a una curva logarítmica, tal como es posible verlo en la Figura 38). Posteriormente, el “codo” de la curva se selecciona como la cantidad óptima.

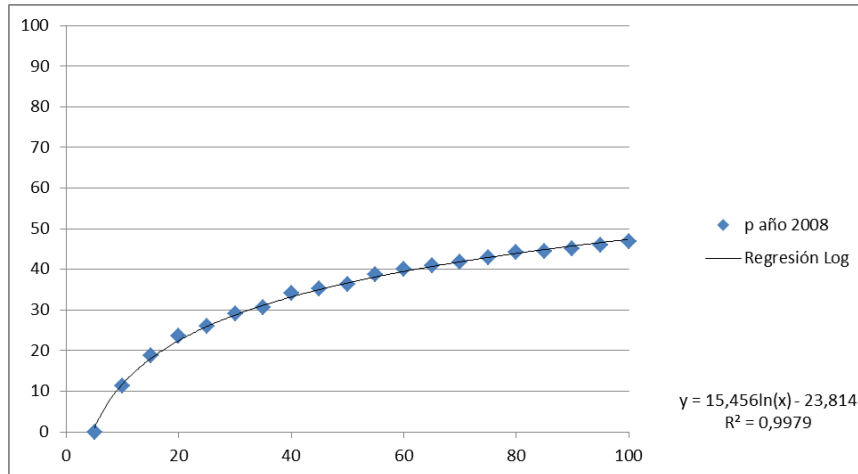


Figura 38. Estimación de la cantidad de tópicos para el año 2008 en la Comunidad I.

En la Figura 39 es posible ver la cantidad óptima de tópicos para cada periodo analizado.

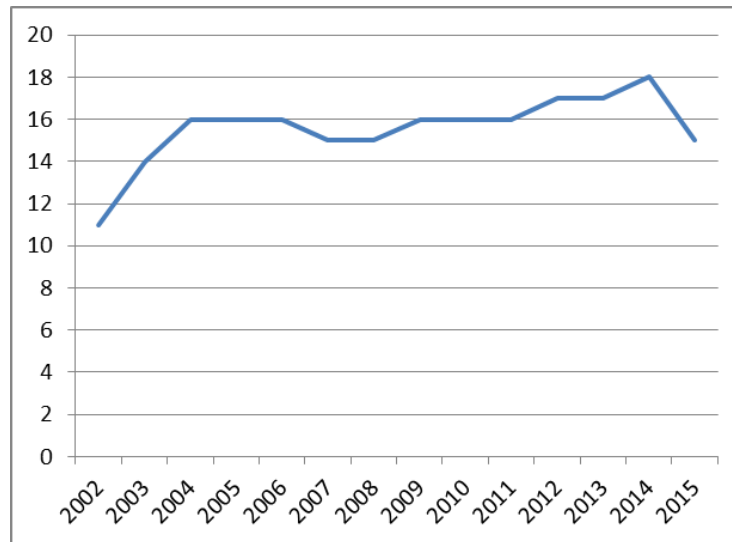


Figura 39. Cantidad de tópicos para cada período para la Comunidad I.

(c) Clasificación de tópicos

Posteriormente, los administradores revisan los tópicos para cada periodo, y los clasifican en tres categorías: (a) tópico altamente relacionado con la comunidad, (b) tópico levemente relacionado con la comunidad, y (c) tópico no relacionado con la comunidad. En la Tablas 9 y 10 es posible ver ejemplos de dichas clasificaciones. En el primer ejemplo (Tabla 9) se encuentran dos tópicos, correspondientes al año 2006, formados por palabras propias de la comunidad, como por ejemplo “cable”, “ruido”, “guitarra”, “volumen”, entre otras. En el segundo ejemplo (Tabla 10) se encuentran dos tópicos, correspondientes al año 2005, formados por palabras que no tienen generalmente relación con la comunidad, y que son utilizadas más bien para expresar ideas generales o como elementos para formar otras oraciones, como por ejemplo, “gracias”, “saludos”, “saber”, entre otras.

Tabla 9. Ejemplos de tópicos con alta relación con la Comunidad I.

Año 2006 / Tópico 5		Año 2006 / Tópico 6	
cable	0.02614838890648233	placa	0.0185159180873805
ruido	0.02518864483823386	papel	0.012016382339933889
cables	0.01703082025812189	queda	0.009737324350569494
tierra	0.014204907168279182	caja	0.009568505240246206
problema	0.0133518013298361	placas	0.009315276574761272
tiene	0.011485632308241858	aluminio	0.007964723692174965
guitarra	0.011219036733728394	son	0.007373856806043454
gracias	0.008872995678009919	agua	0.007205037695720166
volumen	0.008286485414080301	después	0.006867399475073588
jack	0.008233166299177608	pintura	0.006867399475073588

Tabla 10. Ejemplos de tópicos sin relación con la Comunidad I.

Año 2005 / Tópico 1		Año 2005 / Tópico 2	
alguien	0.04321418400426131	hacer	0.018079768605328794
gracias	0.03086658751439567	bueno	0.01741729758843578
hola	0.02690603769689159	bien	0.016066875900153876
saludos	0.01991683213659028	saludos	0.015175087992797898
tengo	0.015102046083938262	mejor	0.014424654351846483
saber	0.014247809848790323	ahora	0.013034797015143559
necesito	0.013548889292760193	así	0.012717517500622185
sabe	0.013548889292760193	estoy	0.010919985691985101
algún	0.012772310897171157	después	0.010028197784629125

venden	0.01215104818069993	parece	0.009544084349207308
--------	---------------------	--------	----------------------

Un resumen de la clasificación de los tópicos, hecha por los administradores de la comunidad, se encuentra en la Tabla 11.

Tabla 11. Resumen de la clasificación de tópicos para la Comunidad I.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
TOTAL	11	14	16	16	16	15	15	16	16	16	17	17	18	15
SI	6	10	10	10	10	9	10	10	11	11	13	14	12	11
NEUTRO	1	1	2	2	2	3	2	3	3	4	1	0	3	3
NO	4	3	4	4	4	3	3	3	2	1	3	3	3	1
% SI	55%	71%	63%	63%	63%	60%	67%	63%	69%	69%	76%	82%	67%	73%
% NEUTRO	9%	7%	13%	13%	13%	20%	13%	19%	19%	25%	6%	0%	17%	20%
% NO	36%	21%	25%	25%	25%	20%	20%	19%	13%	6%	18%	18%	17%	7%

A continuación se analiza para cada documento (que corresponde a un hilo de discusión) a cuál tópico corresponde, utilizando uno de los archivos de salida del mismo algoritmo (model-final.theta). En dicho archivo queda registrada la probabilidad con la cual un documento puede ser clasificado en cada tópico. Con esa información y con la clasificación hecha anteriormente sobre la utilidad de cada tópico, se clasifica cada documento de acuerdo a si está relacionado con la comunidad virtual o no. Finalmente se determina el total de documentos que cumple el propósito de la comunidad (ver Figura 40).

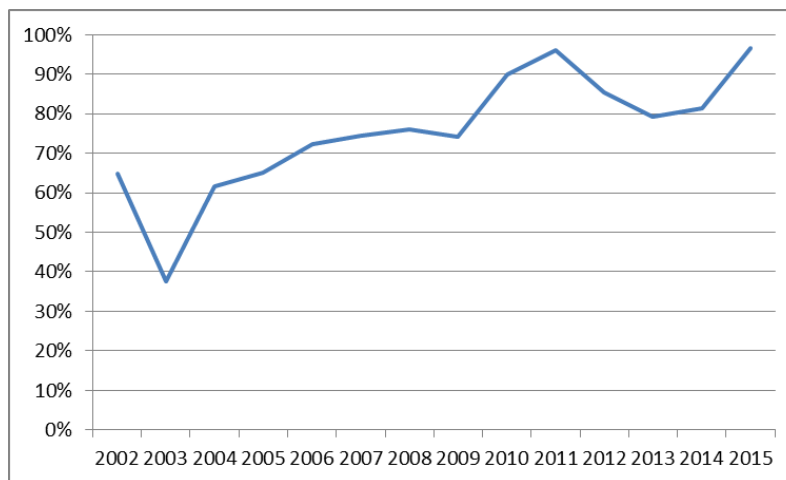


Figura 40. Cumplimiento de propósito para la Comunidad I.

(c) Comparación versus conteo simple

Al igual que la sección anterior, es necesario revisar si al utilizar estas técnicas de evaluación se obtiene un resultado diferente que al haber hecho solamente un conteo de los mensajes que pertenecen a los temas relacionados con la comunidad. Para ello se utiliza como referencia aquellos mensajes que no pertenecen al foro general de la comunidad, y se comparan con el mejor resultado obtenido (por medio de la técnica PYTM) (ver Figura 41).

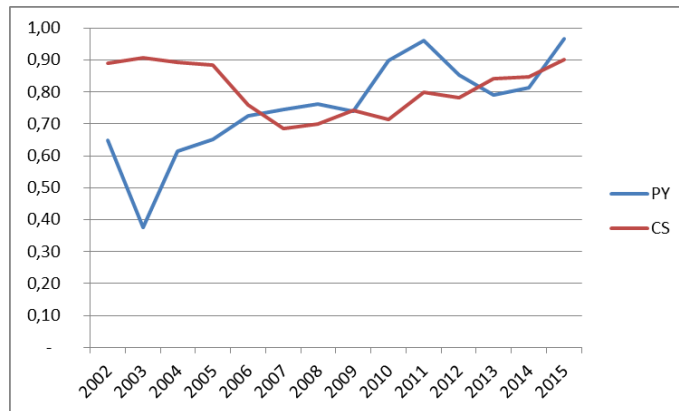


Figura 41. Comparación de resultado obtenido con PYTM y con conteo simple para la Comunidad I.

Tal como es posible ver en la Figura 42, no existe correlación entre los resultados obtenidos con la técnica de PYTM y conteo simple. En efecto, el índice correlación de Pearson es $-0,36$, lo cual indica que no se encuentran correlacionados.

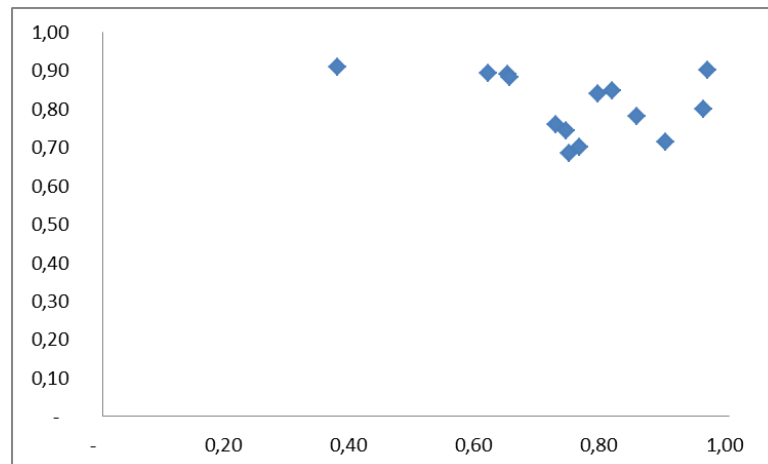


Figura 42. Cruce entre los resultados obtenidos con PYTM y conteo simple para la Comunidad I.

6.4.4 Preparación de Resultados

Los resultados son presentados de dos formas: (a) en forma temporal, para poder comprender cómo fue evolucionando a través del tiempo el cumplimiento del propósito, y (b) como un semáforo, para poder indicar en un periodo determinado el nivel de salud de la comunidad, respecto de la evaluación de este aspecto social (ver Figura 43). Según la definición de los administradores de la comunidad se utilizan los siguientes rangos: Rojo si < 70%, Amarillo entre 70% y 80 %, Verde si > 80 %.

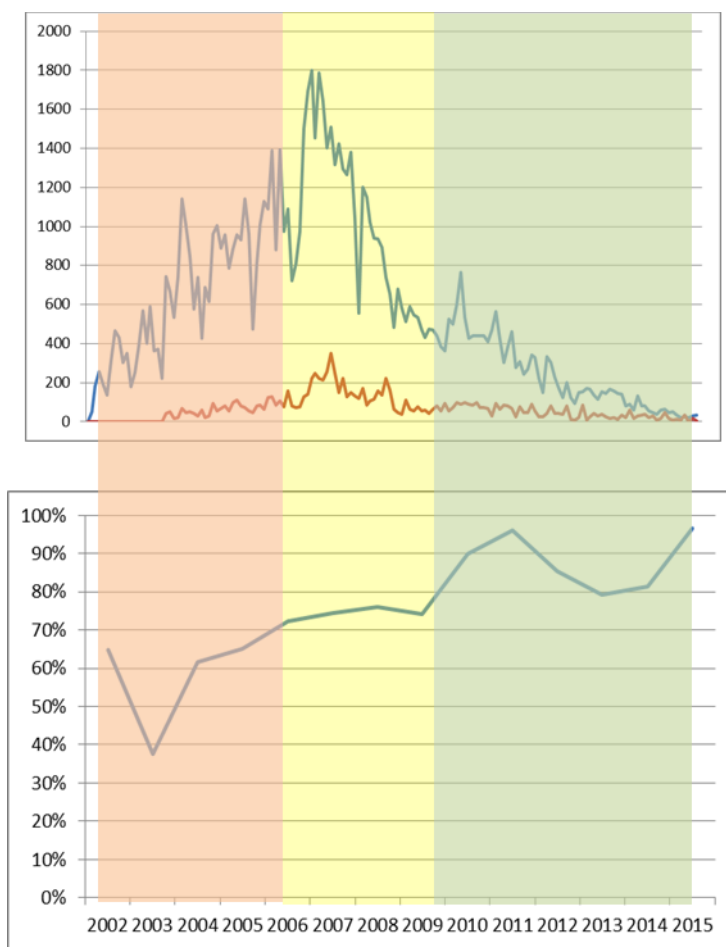


Figura 43. Resumen cumplimiento del propósito para la Comunidad I.

Es posible observar que a medida que va pasando el tiempo, los miembros de la comunidad se han ido alineando más al propósito de la comunidad.

6.4.5 Resultado de la Encuesta

La encuesta para medir el indicador de *claridad* del propósito, fue respondida por 9 usuarios (los mismos miembros que respondieron la encuesta de *identidad*), cuyos resultados se encuentran en la Tabla 12.

Tabla 12. Resultado de la Encuesta de Claridad del Propósito para la Comunidad I.

Pregunta	Muy en desacuerdo			Muy de Acuerdo		
	1	2	3	4	5	PROMEDIO
1	1		1	3	4	4,0
2	2		6	1		2,7
						3,3

Tal como se puede ver, en general la evaluación es regular. La nota más baja se obtiene con la pregunta 2: “¿Sientes que los miembros de la comunidad tienen claro cuál es el objetivo de la comunidad?”.

6.5 Evaluación de la Estrategia Propuesta Políticas

A continuación, para cada una de las comunidades en estudio, se muestra la aplicación de la estrategia propuesta para el aspecto social de *políticas*.

6.5.1 Preparación de los Datos

Para poder evaluar este indicador, la información se extrae directamente desde dos conceptos: (a) relacionados con Moderación, específicamente aquellos relacionados con Moderación de Contextos, para poder determinar la moderación explícita (Tabla 13), y (b) desde las Interacciones, para determinar en cuales Contextos los administradores intervinieron sin necesariamente bloquear la participación de algún miembro (Tabla 14).

Tabla 13. Resumen de las actividades de moderación realizada por los administradores de la Comunidad I.

Tipo de Moderación	Total	Tipo
Moderación Persona	16	usuarios bloqueados
Moderación Contexto	39	tópicos bloqueados
Moderación Contexto	47	tópicos movidos
Moderación Interacción	1.065	posts editados

Tabla 14. Resumen de la actividad de los administradores, para la Comunidad I.

admin	# topics	# posts	# foros	año ini	año fin
1	2794	5221	7	2002	2015
36	1043	1695	7	2002	2015
147	12	28	2	2003	2004
612	8	9	2	2000	2004
615	746	1251	7	2004	2011

6.5.2 Aplicación de Técnicas de Evaluación

Se calculan los totales de interacciones y contextos en los cuales participan los administradores de la comunidad (Ver Figuras 44 y 45).

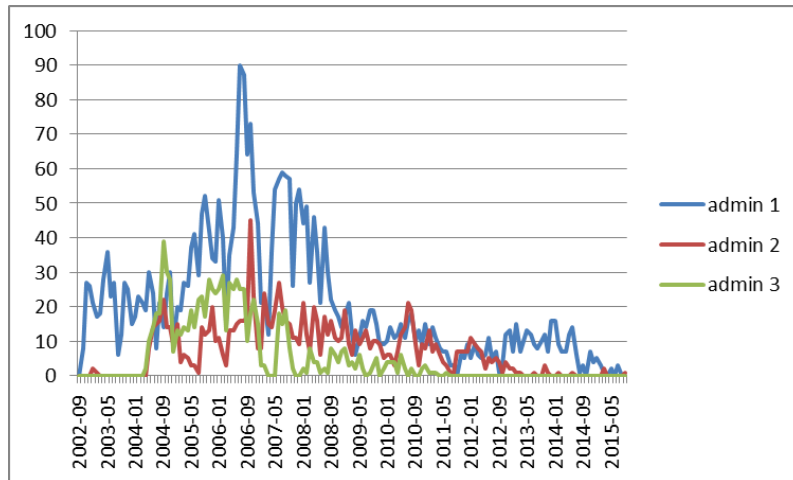


Figura 44. Actividad de los administradores a través de la vida de la Comunidad I.

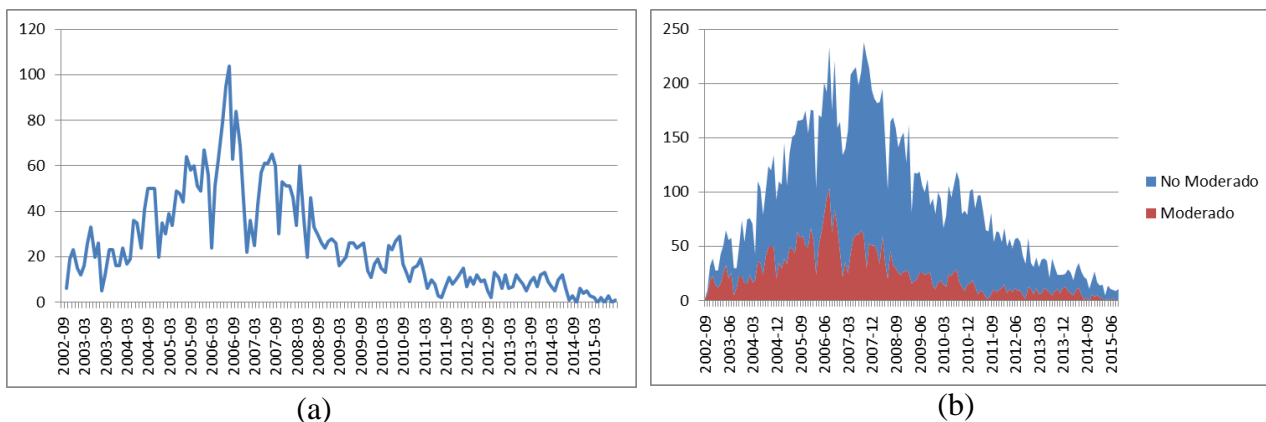


Figura 45. Resumen de intervenciones realizadas por los administradores de la Comunidad I: (a) total de contextos (tópicos) con alguna intervención para la comunidad I (b) comparación entre contextos (tópicos) moderados y no moderados para dicha comunidad.

6.5.3 Generación de Reporte con Resultados

Finalmente, se entrega los resultados, de la misma forma que para los indicadores anteriores. Se calcula finalmente el nivel de moderación de dos formas posibles: (a) considerando las interacciones de los administradores dentro de los contextos (Figura 46), y (b) sin considerar las interacciones, sólo las acciones de moderación explícitas (Figura 47). En ambos casos se despliega el % de contextos que no requieren moderación.

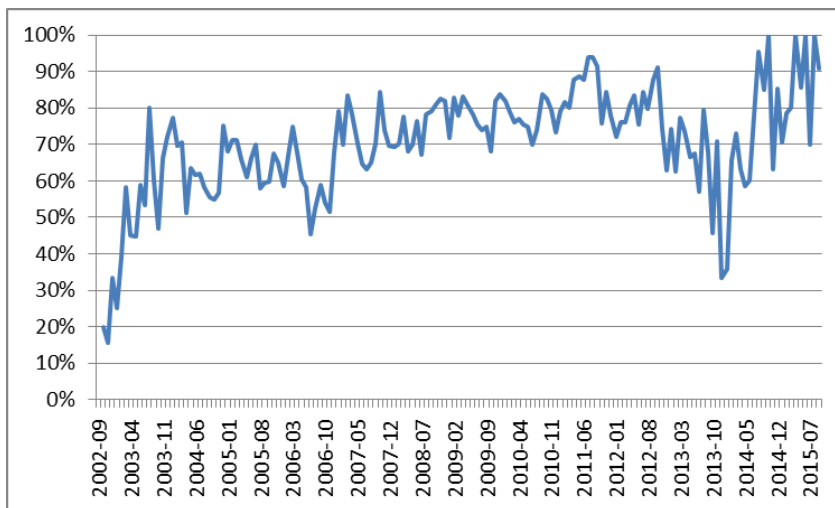


Figura 46. Niveles de Moderación para la Comunidad I, considerando sus interacciones como moderación.

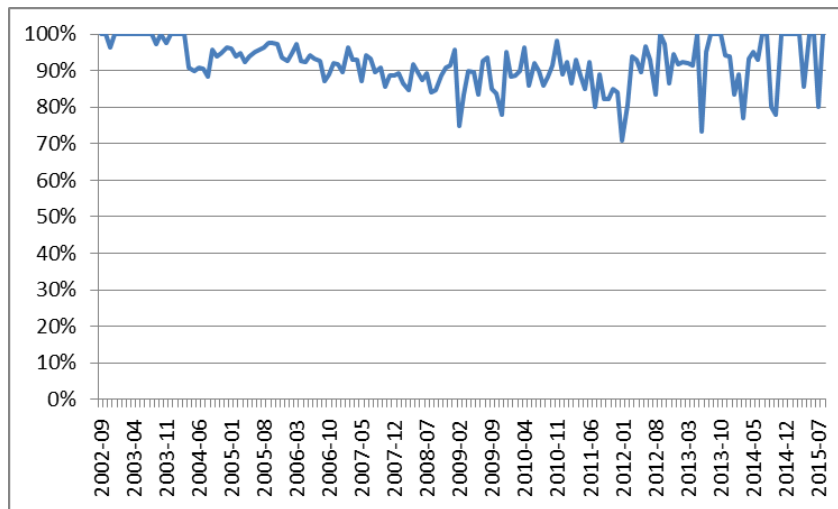


Figura 47. Niveles de Moderación para la Comunidad I, sin considerar sus interacciones como moderación.

El resumen del indicador niveles de moderación se muestra en la Figura 48. Es posible observar que a medida que va pasando el tiempo, los administradores han ido moderando menos, con algunas excepciones. Lo anterior coincide con el hecho de que al final de la vida de la comunidad la participación también ha bajado. Según la definición de los administradores de la comunidad se utilizan los siguientes rangos: Rojo si $< 70\%$, Amarillo entre 70% y 80% , Verde si $> 80\%$.

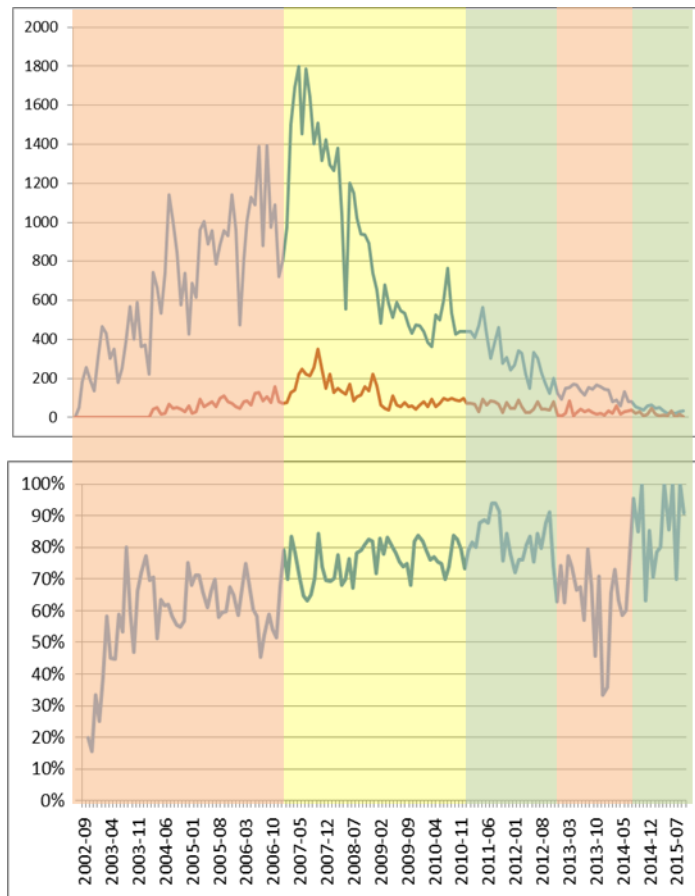


Figura 48. Resumen de nivel de moderación para la Comunidad I.

6.6 Análisis de los Resultados Obtenidos

Finalmente se entrevistó a los dos principales administradores de la comunidad con el objetivo de analizar y evaluar los resultados obtenidos, principalmente en tres momentos específicos de existencia de esta comunidad virtual (ver Figura 49): (a) 2003-09, el cual corresponde a transcurrido un año desde la creación de la comunidad virtual, (b) 2007-05, el cual corresponde al período de mayor actividad de la comunidad, y (c) 2015-07, el cual corresponde al último mes en que fue analizada la comunidad. Los administradores entrevistados identificaron si esperaban los resultados obtenidos y, en caso de que no fuese así, determinaron si les resultaban coherentes de acuerdo al conocimiento previo que ellos poseen acerca de la comunidad virtual.



Figura 49. Resumen de los indicadores para la Comunidad I.

Dentro de los resultados obtenidos más importantes, los cuales eran esperados por los administradores, se encuentran el hecho que:

- La comunidad tuvo un crecimiento explosivo, y durante los últimos años la participación de sus miembros ha decaído considerablemente. La explicación dada por los administradores tiene relación con la existencia de otros foros, de mayor tamaño, que han atraído a los miembros de esta comunidad y causado que dejen de participar. Adicionalmente, la existencia de otro tipo de aplicaciones sociales, como Facebook, ha causado que la interacción entre los miembros de mayor antigüedad se realice por estos nuevos medios.
- En los primeros años de la comunidad virtual la participación y moderación por parte de los administradores fue clave. En los años posteriores, los administradores fueron dejando de participar, dejando que los miembros de la comunidad se moderen entre ellos mismos.

Dentro de los resultados obtenidos más importantes, lo cuales no eran esperados por los administradores, pero que son coherentes completamente con el conocimiento que ellos tienen sobre la comunidad, se encuentra el hecho que:

- El núcleo de la comunidad fue decayendo con los años. La explicación dada por los administradores tiene relación con el hecho de que incluso ellos mismos, quienes fueron parte del núcleo al comienzo de la existencia de la comunidad, dejaron de serlo a medida que transcurrieron los años. Si se analiza la composición del núcleo durante los distintos períodos, es posible confirmar ese hecho, y ver cómo otros miembros de la comunidad fueron alternando su participación en el núcleo de la comunidad.
- El cumplimiento del propósito fue aumentando con el transcurso de los años. La explicación dada por los administradores tiene relación con el hecho de que los miembros de la comunidad se fueron especializando durante el transcurso de los años. Lo mismo ocasionó aparentemente, según los mismos administradores, que la participación haya decaído desde un momento dado de la comunidad.

Capítulo 7 – Estudio de Comunidad II

Corresponde a una comunidad virtual de práctica creada al interior de una empresa, con el objetivo de crear e intercambiar ideas de innovación. Sus miembros pueden escribir sus ideas y el resto de los miembros puede opinar al respecto. Todas las interacciones entre los miembros de esta comunidad son realizadas a través del uso de un sistema creado especialmente para ellos.

La comunidad cuenta con más de 500 miembros en sus más de tres años de experiencia. La tarea de administración es realizada por un funcionario externo contratado especialmente para dicha actividad.

7.1 Análisis de la Comunidad

Se dispone de los datos de la comunidad, desde su formación el año 2013, hasta comienzos del año 2015. La forma de interacción entre los miembros es a través de la publicación de comentarios a ideas escritas por otros miembros. En la Figura 50 es posible visualizar el total de interacciones entre los años 2013 y 2015.

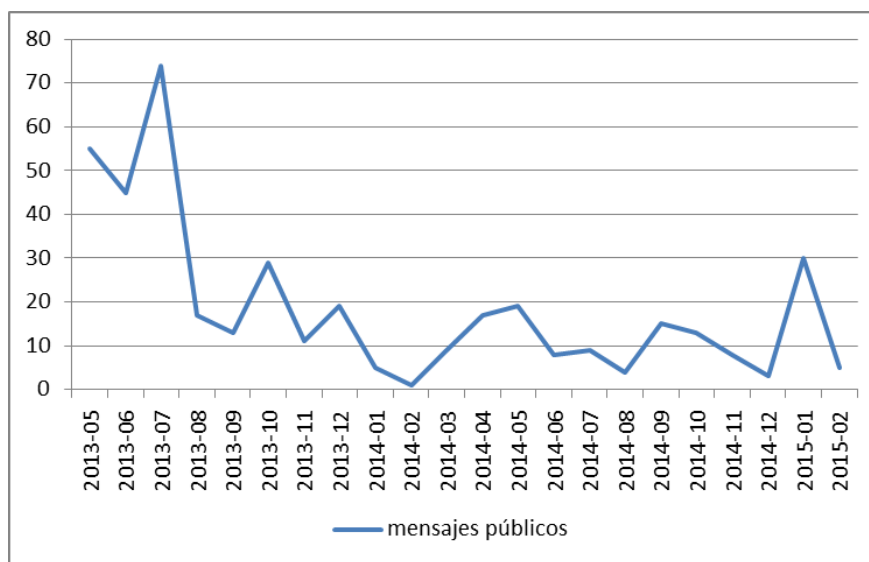


Figura 50. Actividad de la Comunidad II.

En la Figura 51 se muestran los miembros de la comunidad que participan activamente. Es posible observar que la mayor participación se concentra al comienzo de la existencia de la comunidad, y que posteriormente la participación permanece baja durante el resto de los años.

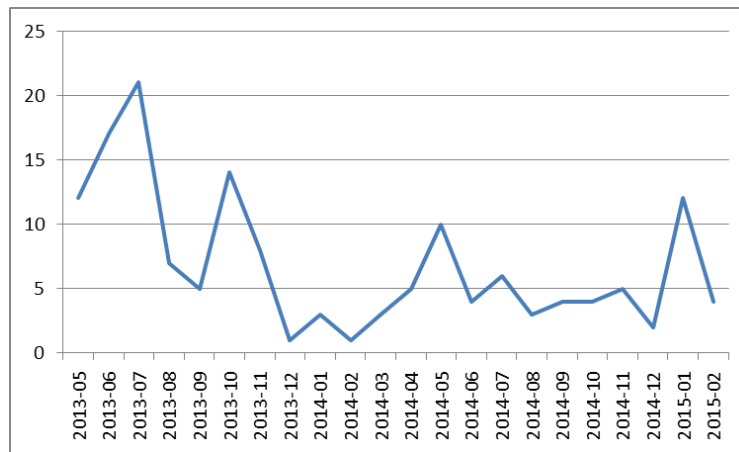


Figura 51. Miembros activos en la Comunidad II.

7.2 Proceso de Extracción de Datos

En la Figura 52 es posible ver el diagrama de clases simplificado del sistema de software utilizado por los miembros de la comunidad virtual. A continuación se explican brevemente cada una de estas clases.

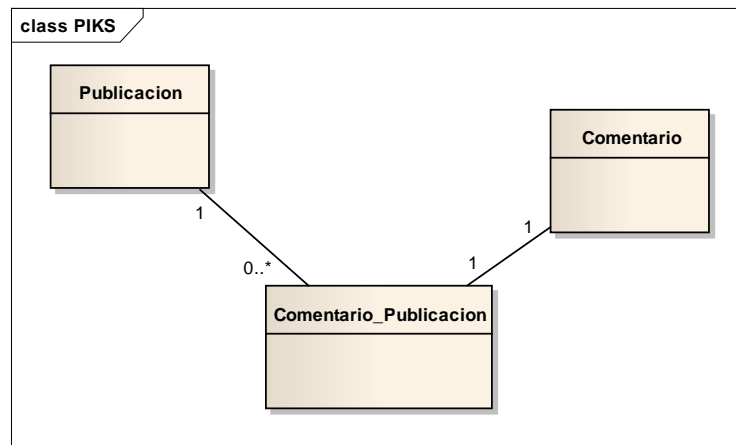


Figura 52. Diagrama de clases simplificado del Foro de Discusión utilizado por la Comunidad II.

Publicación: Corresponde a las ideas de innovación que son publicadas por los miembros de la comunidad.

Comentario: Corresponde a los comentarios de los miembros acerca de alguna idea de innovación.

Comentario_Publicacion: Establece la relación entre un comentario publicado y la publicación a la que pertenece.

En la siguiente tabla se muestra la correspondencia entre los conceptos de esta comunidad, y el modelo social propuesto. En base a esta información se crea el proceso que extrae los datos de esta comunidad virtual hacia el modelo de datos común, en base al cual se aplican las diferentes técnicas de evaluación.

Tabla 15. Correspondencia entre conceptos de la Comunidad II y el modelo social propuesto.

Concepto Social	Comunidad Virtual II
Personas	Se obtienen los usuarios desde las tablas de <i>Publicación</i> y <i>Comentario</i>
Grupos	No Aplica
Roles	No hay
Interacción	<i>Comentario</i> y <i>Comentario_Publicación</i>
Contexto	<i>Publicación</i>
Contenido	<i>Comentario</i>
Tags	No Aplica
Permisos	No hay
Moderación Persona	No hay
Moderación Interacción	No hay
Moderación Contexto	No hay

7.3 Evaluación de la Estrategia Propuesta – Personas

7.3.1 Preparación de los Datos

De la misma forma que la comunidad virtual anterior, se generan las diferentes redes sociales para cada periodo en estudio. El estudio se realiza a nivel mensual, por lo que son 22 periodos (entre el 2013 y 2015), con lo cual se generan 88 redes sociales. Un resumen de las redes sociales es posible visualizar en la Tabla 16.

Tabla 16. Resumen de las redes sociales generadas para la Comunidad II.

	reply_all		creator		reply_prev	
	v	e	v	e	v	e
2013-05	12	66	12	11	12	22
2013-06	17	156	17	16	17	27
2013-07	20	236	0	0	20	31
2013-08	7	26	0	0	7	8
2013-09	5	14	0	0	5	6
2013-10	14	114	14	13	13	20
2013-11	8	43	8	7	8	7
2013-12	0	0	0	0	0	0
2014-01	3	5	0	0	3	2
2014-02	0	0	0	0	0	0
2014-03	3	6	3	2	3	2
2014-04	5	13	4	3	4	5
2014-05	10	46	10	8	10	19
2014-06	4	8	4	3	4	5
2014-07	6	21	5	4	0	0
2014-08	3	6	0	0	3	2
2014-09	4	12	4	4	4	5
2014-10	4	5	4	3	4	3
2014-11	5	20	5	4	5	7
2014-12	0	0	0	0	0	0
2015-01	12	81	12	10	12	14
2015-02	0	0	0	0	0	0

7.3.2 Aplicación de las Técnicas de Evaluación

A continuación se explica cómo se aplican las técnicas de evaluación definidas en la estrategia de evaluación, para la Comunidad II:

(a) Cálculo de periferia

Dado que esta comunidad se encuentra al interior de una organización, para calcular la cantidad de miembros que pertenece a la periferia basta contar el total de miembros de la organización que se encuentran inscritos en la comunidad, y posteriormente restar los miembros activos para un periodo determinado (ver Figura 53).

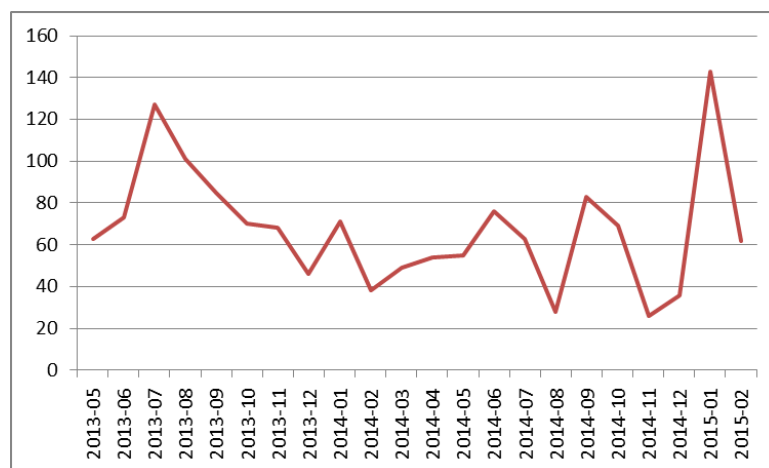


Figura 53. Miembros pertenecientes a la periferia de la Comunidad II.

(b) Cálculo de Miembros Activos y Núcleo

Para cada red social generada (88 en total) se calcula la centralidad de sus nodos en base a la técnicas de *degree*, *betweness centrality*, *closeness centrality*, *hits (authority y hub)*, y *pagerank*.

Para cada resultado obtenido se calcula la cantidad de miembros que pertenecen al núcleo en base a la distribución del índice de centralidad de cada miembro. En la Tabla 17 es posible observar los resultados obtenidos para la técnica *degree*. A diferencia de la comunidad virtual en estudio anterior, no se utiliza la red social *reply_all30*, dado que es casi idéntica a la red *reply_all* para esta comunidad.

Tabla 17. Miembros pertenecientes al núcleo en la Comunidad II, utilizando como técnica de evaluación *degree*.

	degree		
	reply_all	creator	reply_prev
2013-05	6	2	3
2013-06	3	1	3
2013-07	4	0	3
2013-08	2	0	2
2013-09	2	0	2
2013-10	4	2	3
2013-11	3	2	1
2013-12	0	0	0
2014-01	1	0	1
2014-02	0	0	0
2014-03	0	1	1
2014-04	2	1	2
2014-05	5	1	4
2014-06	1	1	2
2014-07	2	2	0
2014-08	0	0	1
2014-09	0	2	2
2014-10	2	1	1
2014-11	0	2	2
2014-12	0	0	0
2015-01	3	1	2
2015-02	0	0	0

7.3.3 Evaluación de los Resultados

A continuación se explica cómo se evalúan los resultados obtenidos, de acuerdo a lo especificado en la estrategia de evaluación:

- (a) cálculo de precisión. Al igual que la comunidad, la técnica que da mejor rendimiento es *degree* con *reply_all*.
- (b) conteo simple. Al igual que con la comunidad anterior, se calculan los niveles de participación utilizando como referencia la cantidad de interacciones de cada miembro (Ver Figura 54).

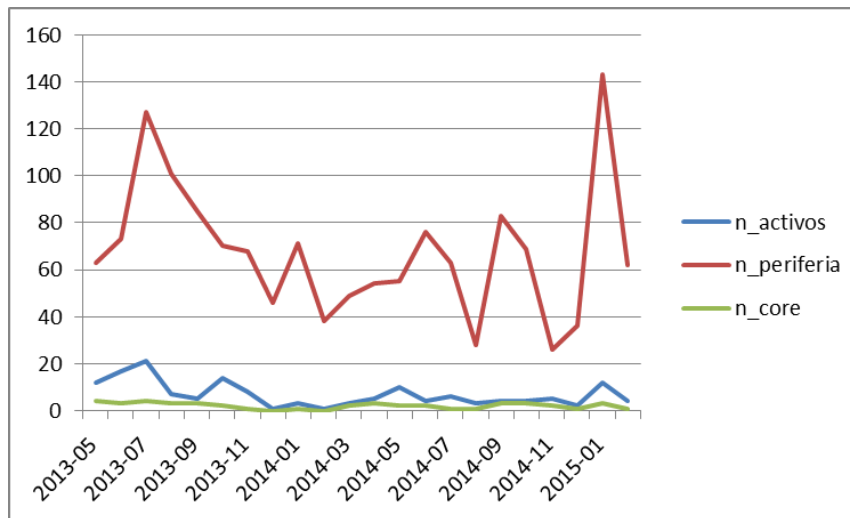


Figura 54. Evolución de los niveles de participación de la Comunidad II, basados en el total de interacciones.

Si se compara el resultado obtenido por conteo simple, con el mejor resultado obtenido con las técnicas de SNA (*degree* sobre la red *reply_all*) (ver Figura 55), es posible visualizar que, en general, la técnica de SNA da mejores resultados, al igual como sucede con la Comunidad I.

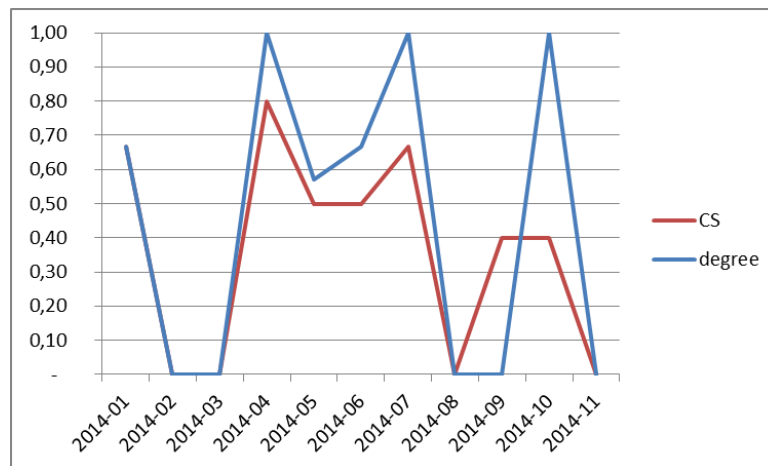


Figura 55. Comparación entre el mejor resultado obtenido con las técnicas SNA y conteo simple en la Comunidad II.

7.3.4 Generación de Reporte

Los resultados son presentados de la misma forma que para la comunidad anterior: (a) en forma temporal, para poder comprender cómo fue evolucionando a través del tiempo los niveles de participación, y (b) como un semáforo, para poder indicar en un periodo determinado el nivel de salud de la comunidad, respecto de la evaluación de este aspecto social. En base a la técnica

degree con la red social *reply_all*, se calcula la evolución temporal de los niveles de participación, contando los miembros que pertenecen a cada nivel de participación, y calculándolo en forma porcentual respecto del total de miembros de la comunidad en un instante de tiempo determinado (Figura 56).

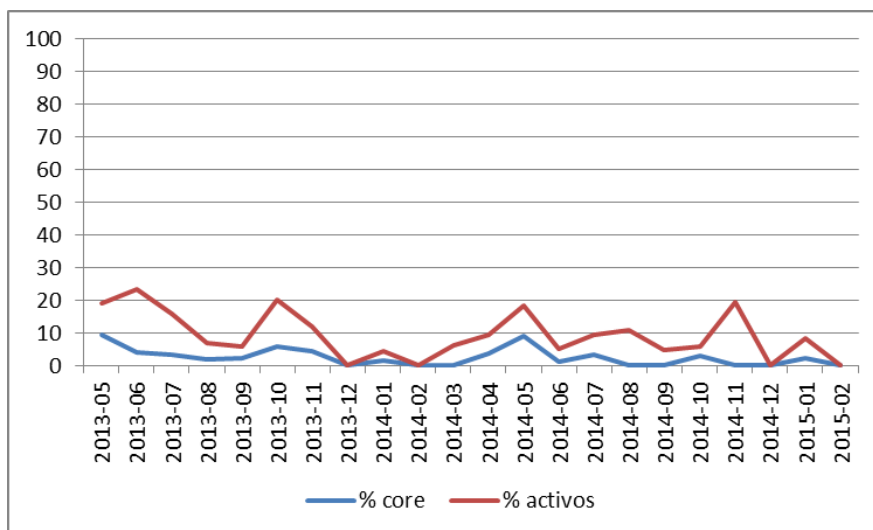


Figura 56. Evolución de los niveles de participación en forma porcentual en la Comunidad II.

Según la definición de los administradores de la comunidad se utilizan los rangos indicados en la Tabla 18.

Tabla 18. Rangos para determinar si niveles de participación son adecuados para la Comunidad II.

Color	Núcleo	Miembros Activos
Rojo	< 10%	< 20%
Amarillo	~ 10 %	20-25 %
Verde	> 10 %	> 25 %

Con la definición anterior, es posible visualizar (Figuras 57 y 58) un resumen de los niveles de participación.

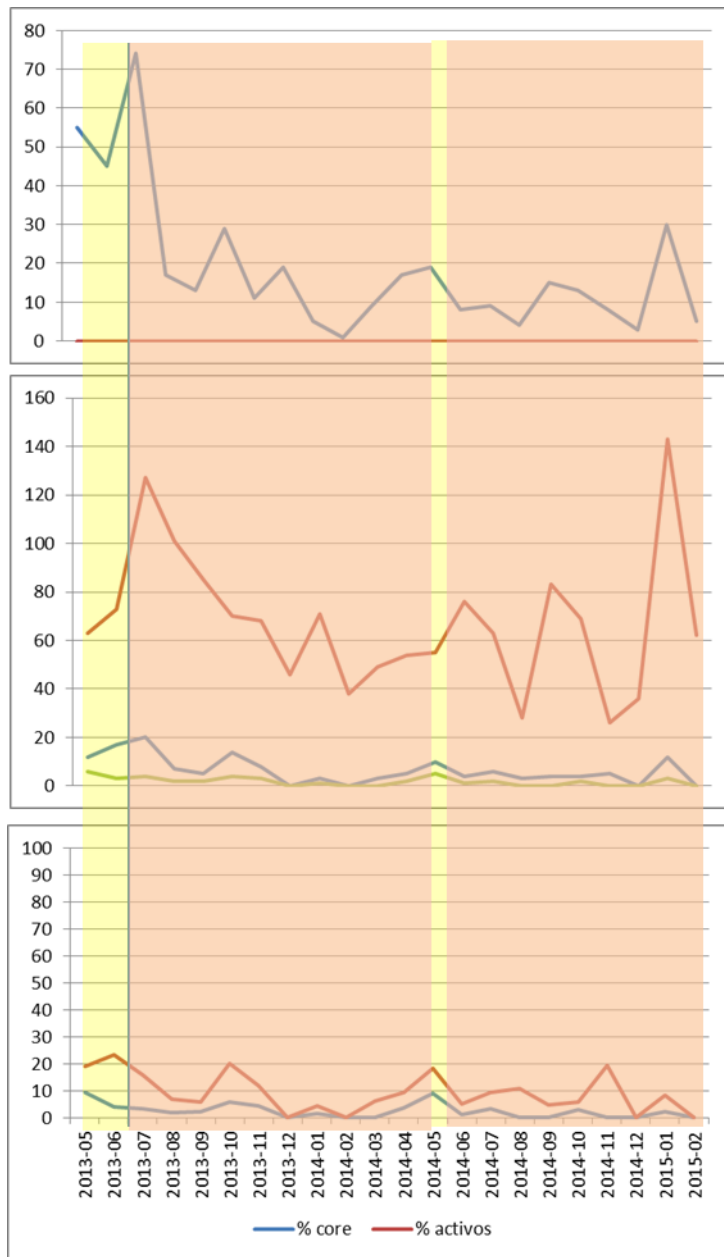


Figura 57. Resumen miembros activos de la Comunidad II.

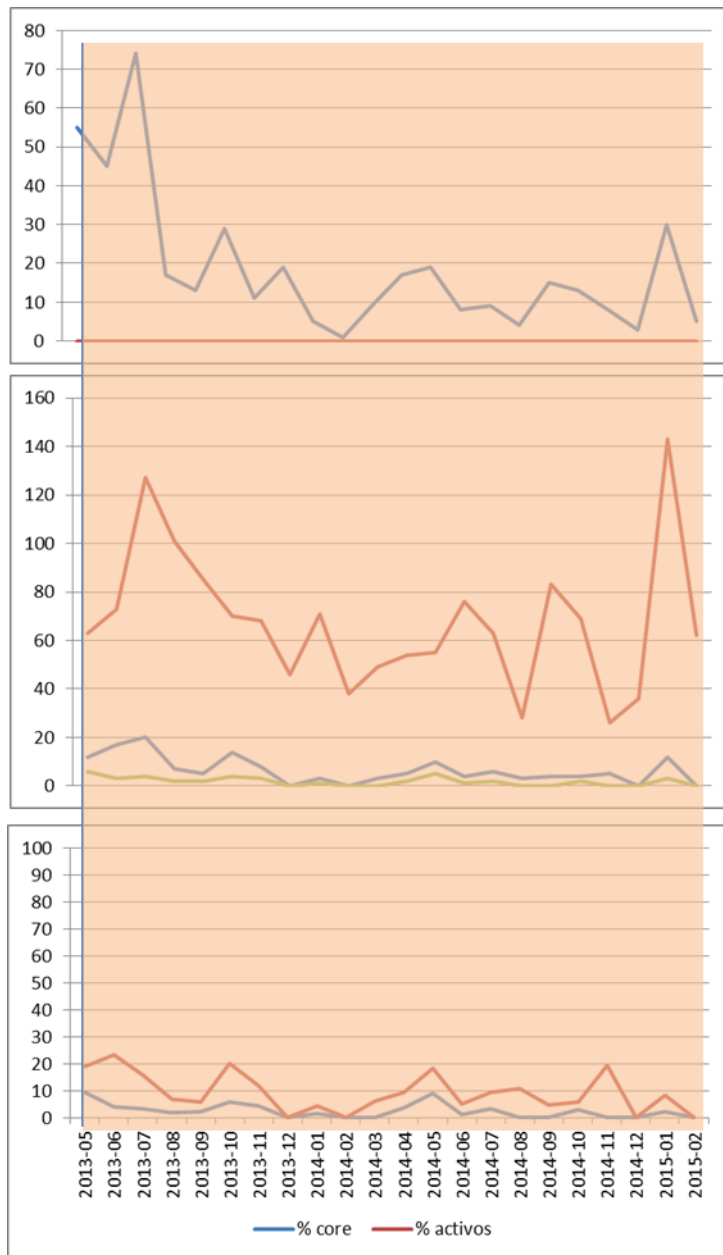


Figura 58. Resumen miembros pertenecientes al núcleo de la Comunidad II.

En estas figuras es posible observar que el nivel de participación del núcleo siempre ha sido bajo. Por otro lado, el nivel de participación de los miembros activos fue “moderado” sólo al comienzo de la comunidad, después también ha sido bajo.

7.4. Evaluación de la Estrategia Propuesta - Propósito

7.4.1 Preparación de los Datos

Al igual que la comunidad I, se generan los archivos necesarios para el procesamiento. En la Tabla 19 se resumen dichos archivos.

Tabla 19. Resumen de los documentos para analizar de la Comunidad II.

Año	total docs	total palabras	promedio palabras / doc	palabras únicas	promedio palabras únicas / doc
2013	202	2725	13,49	1290	6,39
2014	77	1682	21,84	924	12,00
2015	26	707	27,19	432	16,62

7.4.2 Aplicación de Técnicas de Evaluación

Al igual que en la comunidad I, los tiempos de procesamiento para PYTM son mayores que lo de LDA, tal como se muestra en la Figura 59.

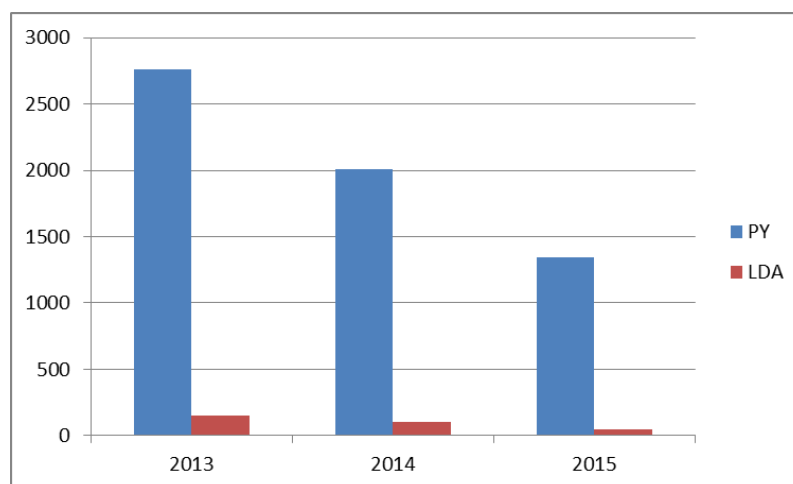


Figura 59. Tiempos de procesamiento para las técnicas de PYTM y LDA para la Comunidad II.

7.4.3 Evaluación de los Resultados

(a) Selección de técnica de evaluación

Al igual que la comunidad I, la técnica PYTM genera mejor rendimiento que LDA (ver Figura 60).

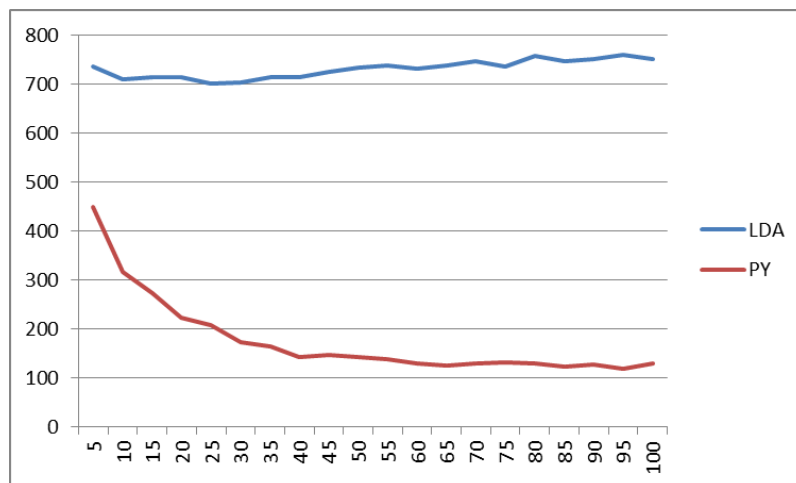


Figura 60. Rendimiento considerando PYTM y LDA, para la Comunidad II, para el año 2013.

(b) Selección de cantidad de tópicos

Para estimar la cantidad óptima de tópicos para cada periodo en estudio, se realiza el mismo procedimiento que para la Comunidad I (Figuras 61). Lo anterior se basa en el cálculo de la regresión sobre los valores obtenidos del índice de *perplelity* (Figura 62) para cada período en estudio.

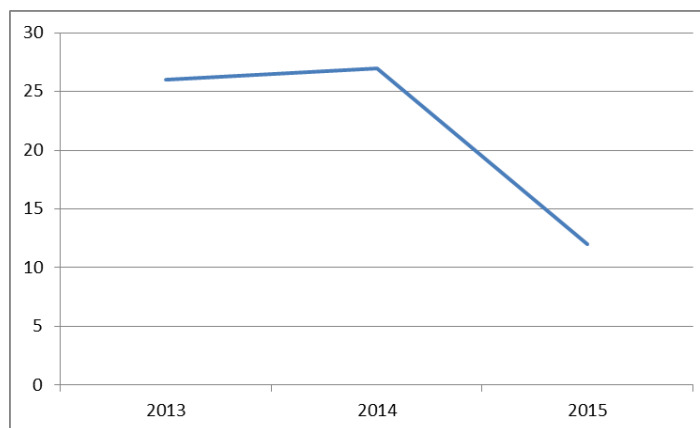


Figura 61. Cantidad de tópicos para cada período para la Comunidad II.

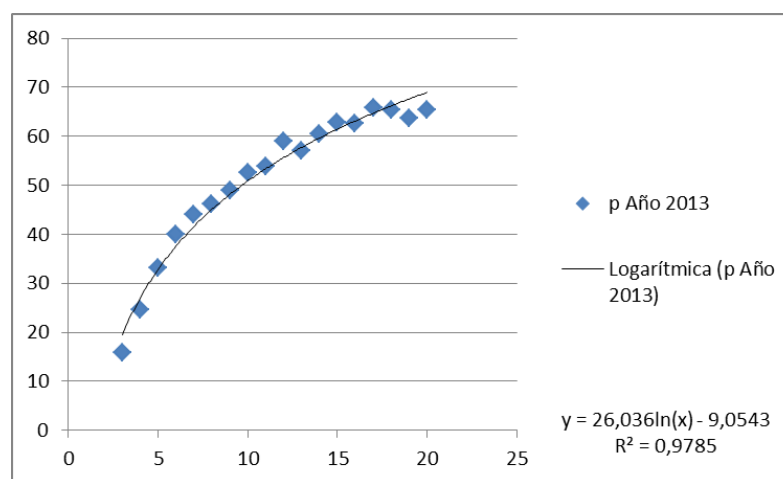


Figura 62. Estimación de la cantidad de tópicos para el año 2013 en la Comunidad II.

(c) Clasificación de tópicos

Al igual que en la comunidad I, el administrador clasifica los tópicos en 3 categorías: (a) tópico altamente relacionado con la comunidad, (b) tópico levemente relacionado con la comunidad, y (c) tópico no relacionado con la comunidad. Algunos ejemplos de los tópicos clasificados se encuentran en la Tablas 20 y 21.

Tabla 20. Ejemplo de tópicos con alta relación con la Comunidad II.

Año 2013 / tópico 1		Año 2013 / tópico 17	
agua	0.08923083346820194	solares	0.07970489060345959
dulce	0.040772683688466274	paneles	0.06261087350944249
caligus	0.031962111001241605	prueba	0.04551685641542539
centros	0.031962111001241605	generadores	0.036969847868416846
sistema	0.02755682465762927	pontones	0.028422839321408298
peces	0.018746251970404602	escaleras	0.028422839321408298
salinidad	0.00993567928317993	acentos	0.028422839321408298
temperatura	0.00993567928317993	implementación	0.028422839321408298
carga	0.00993567928317993	cod	0.028422839321408298
circuito	0.00993567928317993	instalación	0.01132882227391205

Tabla 21. Ejemplo de tópicos sin relación con la Comunidad II.

Año 2015 / tópico 18		Año 2015 / tópico 3	
asdasdasd	0.32328188720101	favor	0.03190965708869091
asdasd	0.26445835778924526	claudia	0.022294272473306294
asdasdasdasd	0.08798776955395118	muñoz	0.022294272473306294
asas	0.08798776955395118	carla	0.015122094639124064
asdasdasdasdasd	0.02916424014218648	actividades	0.01267888785792168
asdas	0.02916424014218648	personal	0.01267888785792168
4x16br	0.00013848695378791	criticas	0.01267888785792168
administrador	0.00013848695378791	gracias	0.01267888785792168
modelo	0.00013848695378791	externas	0.01267888785792168
cbicosbr	0.00013848695378791	auditorias	0.01267888785792168

Un resumen del total de tópicos clasificados se encuentra en la Tabla 22.

Tabla 22. Resumen de la clasificación de tópicos para la Comunidad II.

	2013	2014	2015
TOTAL	26	27	12
SI	21	17	5
NEUTRO	3	1	1
NO	2	9	6
% SI	81%	63%	42%
% NEUTRO	12%	4%	8%
% NO	8%	33%	50%

Al igual que en la comunidad I, finalmente se clasifican todas las interacciones en los diversos tópicos, y se determina el nivel de cumplimiento del propósito (ver Figura 63).

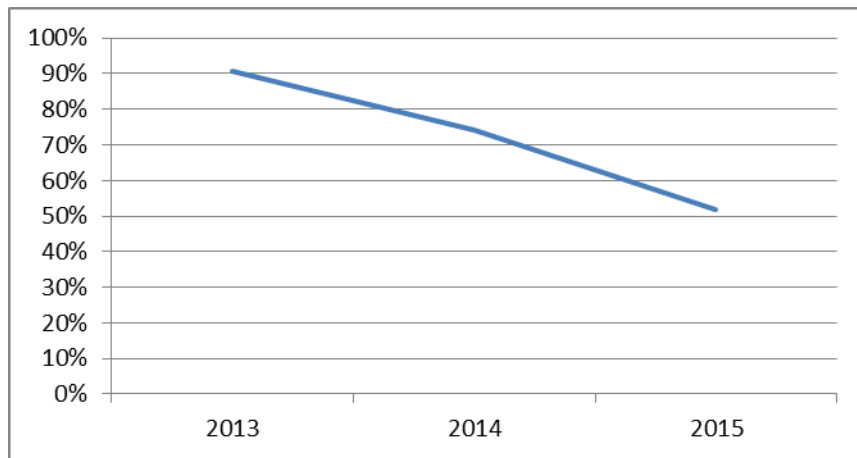


Figura 63. Cumplimiento de propósito para la Comunidad II.

(d) Comparación versus conteo simple

Al igual que con la Comunidad I, se compara los resultados obtenidos con el hecho de solamente haber contado los mensajes que pertenecen a la categoría “General” (Figura 64).

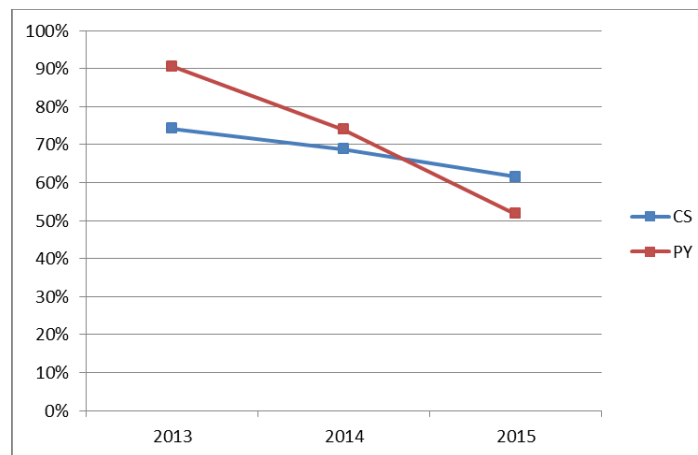


Figura 64. Comparación de resultado obtenido con PYTM y con conteo simple para la Comunidad II.

7.4.4 Generación de Reporte con Resultados

Al igual que con la comunidad II, se presentan los resultados a los administradores (ver Figura 65). Según la definición de los administradores de la comunidad se utilizan los siguientes rangos: Rojo si < 70%, Amarillo entre 70% y 80 %, Verde si > 80 %.

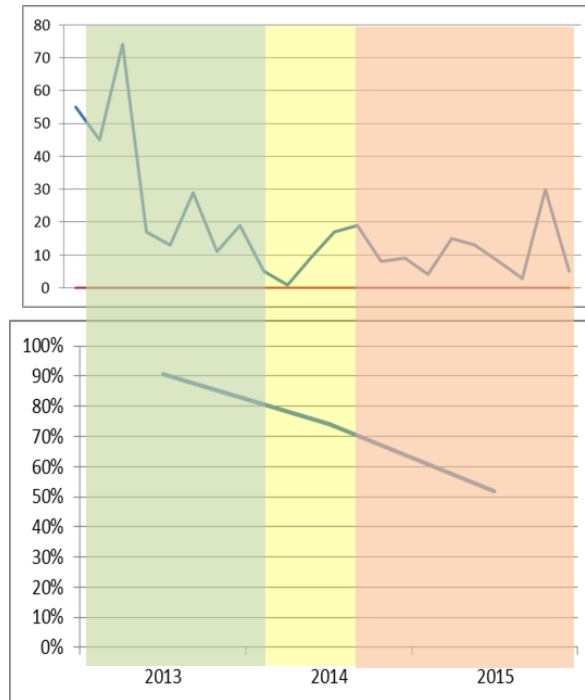


Figura 65. Resumen cumplimiento del propósito para la Comunidad II.

Es posible observar que a medida que va pasando el tiempo, los miembros de la comunidad se han ido alineando menos al propósito de la comunidad.

7.5 Evaluación de la Estrategia Propuesta Políticas

No se aplica el proceso de evaluación, dado que el administrador no interviene en la comunidad.

7.6 Análisis de los Resultados Obtenidos

Finalmente se entrevistó al administrador de la comunidad, con el objetivo de analizar y evaluar los resultados obtenidos, principalmente en dos momentos específicos de existencia de esta comunidad virtual (ver Figura 66): (a) 2013-07, el cual corresponde al período de mayor actividad de la comunidad, y (c) 2015-01, el cual corresponde al último mes que fue analizado.



Figura 66. Resumen de los indicadores para la Comunidad II.

Dentro de los resultados obtenidos más importantes, los cuales eran esperados por el administrador, se encuentran el hecho que:

- Los niveles de participación, tanto del núcleo, como de los miembros activos, es bajo. La explicación dada por el administrador tiene relación con el hecho de que la participación de los miembros de la comunidad es más bien ocasional, y que aún no logran que un grupo de miembros participe en forma sostenida.
- El cumplimiento del propósito ha ido decayendo con el transcurrir del tiempo, lo cual se explica por el hecho de que principalmente al comienzo de la existencia de la comunidad el sistema de software fue utilizado efectivamente para el objetivo para la cual fue creado.

Dentro de los resultados obtenidos más importantes, los cuales no eran esperados por el administrador, pero que son coherentes completamente con el conocimiento que él tiene sobre la comunidad, se encuentra el hecho que:

- La cantidad de tópicos de los cuales hablan los miembros de la comunidad es alta, en comparación con la comunidad anterior. Lo anterior se explica por el objetivo de la comunidad, el cual corresponde a compartir ideas.

Capítulo 8: Conclusiones y Trabajo a Futuro

Internet ha cambiado sustancialmente la forma en la cual las personas interactúan entre sí. Cada día es posible ver más frecuentemente estructuras sociales que emergen, crecen y se desarrollan en el tiempo. Sin embargo, la falta de herramientas de evaluación específicas ocasiona que los administradores de dichas estructuras no tengan información exacta de lo que sucede al interior de las mismas, lo cual ocasiona inevitablemente la desaparición de muchas de ellas. Este es el caso de las comunidades virtuales, las cuales requieren de la atención constante por parte de sus administradores, quienes generalmente no cuentan con una visión completa de lo que sucede al interior de la comunidad, principalmente cuando la cantidad de miembros y de interacciones aumenta con el transcurrir de los años.

El presente trabajo ha mostrado una estrategia para evaluar algunos aspectos sociales que se desarrollan al interior de una comunidad virtual. Se ha mostrado, a través de experimentación con comunidades virtuales existentes, que es posible contribuir con las tareas de los administradores al otorgarles información útil y veraz acerca del funcionamiento y estructura de la comunidad que administran. Si bien existe una limitación en el alcance de los resultados obtenidos, al haber aplicado la estrategia propuesta en dos comunidades virtuales reales en un escenario donde existen innumerables estructuras sociales, cada una con sus propias características y necesidades, el presente trabajo tiene el valor de generar una estrategia de procesamiento que es posible replicar en otras comunidades virtuales. El hecho anterior se basa en que la estrategia propuesta desarrolla con claridad el “cómo” evaluar una comunidad virtual, en un área de investigación predominada por teorías psicológicas y sociológicas que sólo especifican el “qué” evaluar. Dado que la estrategia propuesta basa su análisis en los aspectos sociales definidos por autores del área de sociología, es que se encuentra garantizado, al menos teóricamente, que la estrategia planteada es posible aplicarla a otras comunidades virtuales.

Por lo tanto, es posible concluir que la importancia de los resultados en este trabajo se encuentra principalmente en dos aspectos: (i) debido a la naturaleza social de las comunidades virtuales, es fundamental aplicar conceptos y teorías de la sociología en la evaluación de sistemas de software, de la misma forma que se aplican conceptos de psicología cognitiva en el diseño de interfaces y en las evaluaciones de usabilidad, y (ii) debido a la naturaleza cambiante de las comunidades virtuales, contar con una estrategia de evaluación es la base para poder construir a futuro una metodología para el desarrollo de soluciones para este tipo de estructuras sociales.

Desde el punto de vista de las comunidades virtuales estudiadas, es posible concluir que: (a) a los administradores de las comunidades virtuales (en estudio) les resulta útil tener información acerca del funcionamiento de la comunidad desde un punto de vista sociológico, ya que les permite encontrar explicación a fenómenos que suceden al interior de la comunidad, (b) el uso de técnicas avanzadas de análisis de datos, tales como los modelos de tópicos, permiten generar información útil y veraz a los administradores acerca del funcionamiento de la comunidad virtual, (c) los administradores no tienen la posibilidad de obtener los mismo resultados que el presente trabajo si siguen utilizando las técnicas habituales que ellos manejan, las cuales consisten

básicamente en contar las acciones efectuadas por los miembros de la comunidad, y (d) los resultados obtenidos permiten a los administradores tomar conciencia de los aspectos sociales que desarrollan los miembros de una comunidad virtual al interactuar a través de un sistema de software, y que estos aspectos pueden influir en el éxito o fracaso de una comunidad virtual, al igual que los aspectos técnicos o de usabilidad de una aplicación.

Sin embargo, los resultados obtenidos en este trabajo debiesen ser sólo un punto de partida. Aún es posible mejorar la estrategia propuesta en la medida que (a) se siga utilizando con otras comunidades virtuales, de tal forma de poder determinar si existen patrones de comportamiento que se repitan entre las diferentes comunidades, y poder determinar cómo los aspectos culturales influyen en el desarrollo de los aspectos sociales en estudio; (b) los sistemas de software utilizados registren también las visualizaciones de contenido y no sólo las interacciones, de tal forma de poder comprender mejor los patrones de comportamiento de los miembros de la periferia, para así poder comprender el proceso a través del cual se convierten en miembros activos de la comunidad; y (c) se vayan incorporando más aspectos sociales o indicadores a la estrategia de procesamiento, analizando la relación que poseen con los aspectos sociales ya estudiados.

Bibliografía

- [1] P. Lévy, *Cyberculture*. Minneapolis [u.a.]: Univ. of Minnesota Press, 2001.
- [2] E. Wenger, R. McDermott, and W. Snyder, *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press, 2002.
- [3] A. J. Kim, *Community Building on the Web: Secret Strategies for Successful Online Communities*. 2000.
- [4] F. Henri and B. Pudelko, “Understanding and analysing activity and learning in virtual communities,” *Journal of Computer Assisted Learning*, vol. 19, pp. 474–487, Jan. 2003.
- [5] J. Preece, “Online communities: Usability, sociability, theory and methods,” 2001.
- [6] J. Preece, “Etiquette, Empathy and Trust in Communities of Practice: Stepping-Stones to Social Capital,” *Journal of Universal Computer Science*, Jan. 2004.
- [7] B. Wellman, “An electronic group is virtually a social network,” *Culture of the Internet*, Jan. 1997.
- [8] E. Wenger, N. White, J. Smith, and K. Rowe, “Technology for communities,” *CEFRIO Book Chapter–Jan*, Jan. 2005.
- [9] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, “CRISP-DM 1.0 Step-by-step data mining guide,” Aug. 2000.
- [10] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite, “Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community,” *Annual Review of Sociology*, vol. 22, no. 1, pp. 213–238, 1996.
- [11] B. Wellman, “Computer Networks As Social Networks,” *Science*, vol. 293, pp. 2031–2035, 2001.
- [12] B. Wellman and M. Gulia, “Virtual communities as communities,” *Communities in cyberspace*, Jan. 1999.
- [13] C. Johnson, “A survey of current research on online communities of practice,” *The internet and higher education - Elsevier*, Jan. 2001.
- [14] J. Breslin and S. Decker, “The Future of Social Networks on the Internet: The Need for Semantics,” *IEEE Internet Computing*, vol. 11, no. 6, pp. 86–90, 2007.
- [15] B. Wellman, “Which Types of Ties and Networks Give What Kinds of Social Support?,” *Advances in Group Processes*, vol. 9, pp. 207–235, 1992.
- [16] R. Gross and A. Acquisti, “Information revelation and privacy in online social networks,” Alexandria, VA, USA, 2005, pp. 71–80.
- [17] danah michele boyd, “Friendster and publicly articulated social networking,” Vienna, Austria, 2004, pp. 1279–1282.
- [18] J. Donath and D. Boyd, “Public Displays of Connection,” *BT Technology Journal*, vol. 22, no. 4, pp. 71–82, 2004.
- [19] R. S. Burt, “The Social Capital of Structural Holes,” in *The New Economic Sociology: Developments in an Emerging Field*, New York: Russell Sage Foundation., 2002, pp. 148–90.

- [20] A. Portes, "Social Capital: Its Origins and Applications in Modern Sociology," *Annu. Rev. Sociol.*, vol. 24, no. 1, pp. 1–24, Aug. 1998.
- [21] P. S. Adler and S.-W. Kwon, "Social Capital: Prospects for a New Concept," *The Academy of Management Review*, vol. 27, no. 1, pp. 17–40, Jan. 2002.
- [22] R. P. Putnam, "The Strange Disappearance of Civic America," *The American Prospect*, vol. 24, 1996.
- [23] R. P. Putnam, *Bowling Alone*. New York: Simon & Schuster, 2000.
- [24] B. WELLMAN, A. Q. HAASE, J. WITTE, and K. HAMPTON, "Does the Internet Increase, Decrease, or Supplement Social Capital?," *American Behavioral Scientist*, vol. 45, no. 3, pp. 436–455, Nov. 2001.
- [25] B. Wellman and K. A. Frank, "Network capital in a multi-level world: Getting support from personal communities," in *Social capital: Theory and research*, 2001.
- [26] *Social behavior: its elementary forms : Homans, George Caspar, 1910- : Free Download & Streaming : Internet Archive*. 2010.
- [27] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," 2006, pp. 611–617.
- [28] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," 2007, pp. 29–42.
- [29] Y. Zhu, "Measurement and analysis of an online content voting network: a case study of Digg," Raleigh, North Carolina, USA, 2010, pp. 1039–1048.
- [30] S. Barab, "Designing for virtual communities in the service of learning," *The Information Society*, Jan. 2003.
- [31] B. Wellman and University of Toronto. Centre for Urban and Community Studies, *Studying personal communities in East York / Barry Wellman*. [Toronto] : Centre for Urban and Community Studies, University of Toronto, 1982.
- [32] Q. Jones, "Virtual-Communities, Virtual Settlements & Cyber-Archaeology: A Theoretical Outline," *Journal of Computer Mediated Communication*, Jan. 1997.
- [33] H. Rheingold, "A slice of my life in my virtual community," *Global networks: Computers and international communication*, pp. 57–80, Jan. 1993.
- [34] B. Wellman, "Changing connectivity: A future history of Y2. 03K," *socresonline.org.uk*, Jan. 2000.
- [35] J. Marathe, "Creating community online," *Durlacher Research Ltd*, 1999.
- [36] L. Carotenuto, W. Etienne, M. Fontaine, J. Friedman, H. Newberg, M. Muller, M. Simpson, J. Slusher, and K. Stevenson, "Communityspace: Toward flexible support for voluntary knowledge communities," 1999.
- [37] E. Wenger, "Communities of practice: Learning, meaning, and identity," 1999.
- [38] T. Shummer, "Patterns for building communities in collaborative systems," *Proceedings of the 9th European Conference on Pattern Languages and Programs*, 2004.
- [39] E. Wenger, R. McDermott, and W. Snyder, "Cultivating Communities of Practice: A Guide to Managing Knowledge," *Harvard Business School Press*, 2002.

- [40] G. Probst and S. Borzillo, "Why communities of practice succeed and why they fail," *European Management Journal*, vol. 26, no. 5, pp. 335–347, 2008.
- [41] A. Bourhis, L. Dubé, R. Jacob, and others, "The success of virtual communities of practice: The leadership factor," *The Electronic Journal of Knowledge Management*, vol. 3, no. 1, pp. 23–34, 2005.
- [42] J. Plaskoff, "Intersubjectivity and community building: Learning to learn organizationally," in *Collection*, 2003, pp. 161–184.
- [43] E. D. Mynatt, A. Adler, M. Ito, and V. L. O'Day, "Design for network communities," Atlanta, Georgia, United States, 1997, pp. 210–217.
- [44] C. S. de Souza and J. Preece, "A framework for analyzing and understanding online communities," *Interacting with Computers*, vol. 16, no. 3, pp. 579–610, Jun. 2004.
- [45] H. Saint-Onge and Debra Wallace Ph, *Leveraging Communities of Practice for Strategic Advantage*. Butterworth-Heinemann, 2002.
- [46] L. Garton, C. Haythornthwaite, and B. Wellman, "Studying online social networks," *Journal of Computer-Mediated Communications*, vol. 3, no. 1, 1997.
- [47] P. Selznick, "In search of community," in *Rooted in the land*, 1996.
- [48] E. Lesser and J. Storck, "Communities of practice and organizational performance," *IBM Systems Journal*, vol. 40, no. 4, pp. 831–841, 2001.
- [49] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2012, pp. 850–858.
- [50] S. Siersdorfer, S. Chelaru, J. S. Pedro, I. S. Altingovde, and W. Nejdl, "Analyzing and Mining Comments and Comment Ratings on the Social Web," *ACM Trans. Web*, vol. 8, no. 3, pp. 17:1–17:39, Jul. 2014.
- [51] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4065–4074, Aug. 2013.
- [52] C. M. Chiu, M. H. Hsu, and E. T. G. Wang, "Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories," *Decision Support Systems*, vol. 42, no. 3, pp. 1872–1888, 2006.
- [53] K.-Y. Lin and H.-P. Lu, "Why people use social networking sites: An empirical study integrating network externalities and motivation theory," *Computers in Human Behavior*, vol. 27, no. 3, pp. 1152–1161, May 2011.
- [54] Z. Yan, T. Wang, Y. Chen, and H. Zhang, "Knowledge sharing in online health communities: A social exchange theory perspective," *Information & Management*, vol. 53, no. 5, pp. 643–653, Jul. 2016.
- [55] Z. Zhou, X.-L. Jin, and Y. Fang, "Moderating role of gender in the relationships between perceived benefits and satisfaction in social virtual world continuance," *Decision Support Systems*, vol. 65, pp. 69–79, Sep. 2014.

- [56] D. Cartwright and F. Harary, “Structural balance: a generalization of Heider’s theory,” *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.
- [57] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee, “How Opinions Are Received by Online Communities: A Case Study on Amazon.Com Helpfulness Votes,” in *Proceedings of the 18th International Conference on World Wide Web*, New York, NY, USA, 2009, pp. 141–150.
- [58] N. Sun, P. P.-L. Rau, and L. Ma, “Understanding lurkers in online communities: A literature review,” *Computers in Human Behavior*, vol. 38, pp. 110–117, Sep. 2014.
- [59] H. Zhu, R. E. Kraut, and A. Kittur, “Effectiveness of Shared Leadership in Wikipedia,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 55, no. 6, pp. 1021–1043, Dec. 2013.
- [60] R. West, H. S. Paskov, J. Leskovec, and C. Potts, “Exploiting Social Network Structure for Person-to-Person Sentiment Analysis,” *arXiv:1409.2450 [physics]*, Sep. 2014.
- [61] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, “How Community Feedback Shapes User Behavior,” in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- [62] P. Raeth, S. Smolnik, N. Urbach, and C. Zimmer, “Towards Assessing the Success of Social Software in Corporate Environments,” *AMCIS 2009 Proceedings*, Jan. 2009.
- [63] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, Feb. 2010.
- [64] P. Kim and S. Kim, “Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering,” *Physica A: Statistical Mechanics and its Applications*, vol. 417, pp. 46–56, Jan. 2015.
- [65] J. Gairín-Sallán, D. Rodríguez-Gómez, and C. Armengol-Asparó, “Who exactly is the moderator? A consideration of online knowledge management network moderation in educational organisations,” *Computers & Education*, vol. 55, no. 1, pp. 304–312, Aug. 2010.
- [66] U. Matzat and G. Rooks, “Styles of moderation in online health and support communities: An experimental comparison of their acceptance and effectiveness,” *Computers in Human Behavior*, vol. 36, pp. 65–75, Jul. 2014.
- [67] S. L. Toral, M. R. Martínez-Torres, and F. Barrero, “Analysis of virtual communities supporting OSS projects using social network analysis,” *Information and Software Technology*, vol. 52, no. 3, pp. 296–303, Mar. 2010.
- [68] S. A. Myers and J. Leskovec, “The Bursty Dynamics of the Twitter Information Network,” in *Proceedings of the 23rd International Conference on World Wide Web*, New York, NY, USA, 2014, pp. 913–924.
- [69] C. Lipizzi, L. Iandoli, and J. E. R. Marquez, “Combining structure, content and meaning in online social networks: The analysis of public’s early reaction in social media to newly launched movies,” *Technological Forecasting and Social Change*, vol. 109, pp. 35–49, Aug. 2016.

- [70] A. Euerby and C. M. Burns, "Improving Social Connection Through a Communities-of-Practice-Inspired Cognitive Work Analysis Approach," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 56, no. 2, pp. 361–383, Mar. 2014.
- [71] L. Dubé, A. Bourhis, and R. Jacob, "Towards a typology of virtual communities of practice," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 1, pp. 3–13, 2003.
- [72] A. Hassan Zadeh and R. Sharda, "Modeling brand post popularity dynamics in online social networks," *Decision Support Systems*, vol. 65, pp. 59–68, Sep. 2014.
- [73] X. Li, M. Wang, and T.-P. Liang, "A multi-theoretical kernel-based approach to social network-based recommendation," *Decision Support Systems*, vol. 65, pp. 95–104, Sep. 2014.
- [74] R. Y. K. Lau, C. Li, and S. S. Y. Liao, "Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis," *Decision Support Systems*, vol. 65, pp. 80–94, Sep. 2014.
- [75] S. Amer-Yahia, S. Anjum, A. Ghenai, A. Siddique, S. Abbar, S. Madden, A. Marcus, and M. El-Haddad, "MAQSA: A System for Social Analytics on News," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 2012, pp. 653–656.
- [76] R. Jones, S. Sharkey, J. Smithson, T. Ford, T. Emmens, E. Hewis, B. Sheaves, and C. Owens, "Using Metrics to Describe the Participative Stances of Members Within Discussion Forums," *J Med Internet Res*, vol. 13, no. 1, Jan. 2011.
- [77] J. Lave and E. Wenger, "Situated Learning: Legitimate Peripheral Participation," *books.google.com*, Jan. 1991.
- [78] E. M. Rogers and D. L. Kincaid, "Communication Networks: Toward a New Paradigm for Research," p. 386, 1981.
- [79] B. Wellman and S. D. Berkowitz, "Social Structures: A Network Approach," *Emerald Group Publishing Limited*, vol. 15, p. 528, 1998.
- [80] R. L. Breiger, K. M. Carley, P. Pattison, and N. R. C. (U. S.). C. on H. Factors, "Dynamic Social Network Modeling and Analysis: workshop summary, Volume 2002," p. 379, Jan. 2003.
- [81] S. Wasserman and K. Faust, "Social network analysis: Methods and applications," *books.google.com*, Jan. 1994.
- [82] J. Xu and H. Chen, "CrimeNet explorer: a framework for criminal network knowledge discovery," *ACM Transactions on Information Systems (TOIS)*, Jan. 2005.
- [83] W. Huang, S.-H. Hong, and P. Eades, "How people read sociograms: a questionnaire study," Tokyo, Japan, 2006, pp. 199–206.
- [84] J. Scott, "Social network analysis: a handbook," p. 208, Jan. 2000.
- [85] J. Preece and D. Maloney-Krichmar, "Online Communities: Focusing on Sociability and Usability," *Handbook of Human-Computer Interaction*, Jan. 2003.
- [86] D. Fetterman, "Ethnography: Step by Step," *orton.catie.ac.cr*, Jan. 1998.
- [87] J. Nocera, "Ethnography and Hermeneutics in Cybercultural Research Accessing IRC Virtual Communities," *Journal of Computer-Mediated Communication*, Jan. 2002.

- [88] R. Kozinets, "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities," *Journal of Marketing Research*, Jan. 2002.
- [89] L. Robinson and J. Schulz, "New field sites, new methods: new ethnographic opportunities."
- [90] J. Koh, Y. Kim, B. Butler, and G. Bock, "Encouraging participation in virtual communities," *Communications of the ACM*, Jan. 2007.
- [91] F. Sudweeks and S. J. Simoff, "Complementary explorative data analysis: the reconciliation of quantitative and qualitative principles," *Doing Internet Research: Critical Issues and Methods for Examining the Net*, pp. 29–55, 1999.
- [92] A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimera, "Community analysis in social networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 373–380, 2004.
- [93] K. Ehrlich, C. Lin, and V. Griffiths-Fisher, "Searching for experts in the enterprise: combining text and social network analysis," *Proceedings of the 2007 international ACM conference on Supporting group work*, Jan. 2007.
- [94] M. Graves, A. Constabaris, and D. Brickley, "FOAF: Connecting People on the Semantic Web," *Cataloging & Classification Quarterly*, vol. 43, no. 3–4, pp. 191–202, Apr. 2007.
- [95] J. G. Breslin, S. Decker, A. Harth, and U. Bojars, "SIOC: an approach to connect web-based communities," *International Journal of Web Based Communities*, vol. 2, no. 2, pp. 133–142, Jan. 2006.
- [96] S. A. Ríos, F. Aguilera, F. Bustos, T. Omitola, and N. Shadbolt, "Leveraging social network analysis with topic models and the Semantic Web (extended)," *Web Intelligence and Agent Systems: An International Journal*, vol. 11, no. 4, pp. 303–314, 2013.
- [97] S. A. Ríos, F. Aguilera, F. Bustos, T. Omitola, and N. Shadbolt, "Leveraging social network analysis with topic models and the semantic web," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, 2011, vol. 3, pp. 339–342.
- [98] A. J. Kim, *Community Building on the Web: Secret Strategies for Successful Online Communities*. Addison-Wesley Longman Publishing Co., Inc., 2000.
- [99] J. Plaskoff, "Intersubjectivity and community building: Learning to learn organizationally," in *The Blackwell handbook of organizational learning and knowledge management*, Blackwell Publishing, 2003, pp. 161–184.
- [100] F. Henri and B. Pudelko, "Understanding and analysing activity and learning in virtual communities," 2003. [Online]. Available: <http://hal.archives-ouvertes.fr/hal-00190267/>. [Accessed: 28-Apr-2010].
- [101] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. (Eric) Zhao, "Analyzing patterns of user content generation in online social networks," presented at the the 15th ACM SIGKDD international conference, Paris, France, 2009, p. 369.
- [102] D. W. McMillan, "Sense of community," *J. Community Psychol.*, vol. 24, no. 4, pp. 315–325, Oct. 1996.

- [103] J. E. Puddifoot, “Dimensions of community identity,” *J. Community. Appl. Soc. Psychol.*, vol. 5, no. 5, pp. 357–370, Dec. 1995.
- [104] D. Chavis, “Sense of Community Index.”
- [105] J. L. Nasar and D. A. Julian, “The Psychological Sense of Community in the Neighborhood,” *Journal of the American Planning Association*, vol. 61, no. 2, pp. 178–184, Jun. 1995.
- [106] A. L. Blanchard, “Testing a model of sense of virtual community,” *Computers in Human Behavior*, vol. 24, no. 5, pp. 2107–2123, Sep. 2008.
- [107] D. Chavis, K. Lee, and J. Acosta, “The sense of community (SCI) revised: The reliability and validity of the SCI-2,” in *2nd international community psychology conference, Lisboa, Portugal*, 2008.
- [108] D. Blei, A. Ng, and M. Jordan, “Latent dirichlet allocation,” *The Journal of Machine Learning*, Jan. 2003.
- [109] I. Bíró, D. Siklósi, J. Szabó, and A. A. Benczúr, “Linked latent Dirichlet allocation in web spam filtering,” presented at the the 5th International Workshop, Madrid, Spain, 2009, p. 37.
- [110] S. A. Ríos, J. D. Velásquez, H. Yasuda, and T. Aoki, “Web site off-line structure reconfiguration: A web user browsing analysis,” in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2006, pp. 371–378.
- [111] S. A. Rios, J. D. Velásquez, H. Yasuda, and T. Aoki, “Using a self organizing feature map for extracting representative web pages from a web site,” *International Journal of Computational Intelligence Research (IJCIR)*, vol. 2, pp. 159–167, 2006.
- [112] S. A. Rios, J. D. Velasquez, E. S. Vera, H. Yasuda, and T. Aoki, “Establishing guidelines on how to improve the web site content based on the identification of representative pages,” in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI’05)*, 2005, pp. 284–288.
- [113] G. L’Huillier, S. A. Ríos, H. Alvarez, and F. Aguilera, “Topic-based social network analysis for virtual communities of interests in the Dark Web,” presented at the ACM SIGKDD Workshop, Washington, D.C., 2010, pp. 1–9.
- [114] I. Sato and H. Nakagawa, “Topic Models with Power-law Using Pitman-Yor Process,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2010, pp. 673–682.
- [115] R. A. Schwier, “Shaping the metaphor of community in online learning environments,” 2002.

Anexo A. Publicaciones Relacionadas con la Tesis

1. Ríos, S. A., Aguilera, F., Bustos, F., Omitola, T., & Shadbolt, N. (2013). *Leveraging social network analysis with topic models and the Semantic Web (extended)*. *Web Intelligence and Agent Systems: An International Journal* 11(4), 303-314.
2. Ríos, S. A., Silva, R. A., & Aguilera, F. (2012). *A dissimilarity measure for automate moderation in online social networks*. In *Proceedings of the 4th International Workshop on Web Intelligence & Communities* (p. 3). ACM.
3. Ríos, S. A., Aguilera, F., Bustos, F., Omitola, T., & Shadbolt, N. (2011). *Leveraging social network analysis with topic models and the semantic web*. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on* (Vol. 3, pp. 339-342). IEEE.
4. L'huillier, G., Ríos, S. A., Álvarez, H., & Aguilera, F. (2010). *Topic-based social network analysis for virtual communities of interests in the dark web*. In *ACM SIGKDD Workshop on Intelligence and Security Informatics* (p. 9). ACM.
5. Ríos, S. A., & Aguilera, F. (2010). *Web intelligence on the social web*. In *Advanced Techniques in Web Intelligence-I* (pp. 225-249). Springer Berlin Heidelberg.
6. Álvarez, H., Ríos, S. A., Aguilera, F., Merlo, E., & Guerrero, L. (2010). *Enhancing social network analysis with a concept-based text mining approach to discover key members on a virtual community of practice*. In *Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 591-600). Springer Berlin Heidelberg.
7. Ríos, S. A., Aguilera, F., & Guerrero, L. A. (2009). *Virtual communities of practice's purpose evolution analysis using a concept-based mining approach*. In *Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 480-489). Springer Berlin Heidelberg.

Anexo B. Aspectos Sociales de las estructuras sociales en estudio, desde la perspectiva de diversos autores

Autor	Aspecto Social	Descripción
<i>Redes Sociales</i>		
Garton [46]	<i>Actores</i>	Se refiere a las entidades sociales que son parte de la red, por ejemplo, personas, organizaciones, etc.
	<i>Relaciones</i>	Se refiere a las conexiones existentes entre los diferentes actores. Una relación es caracterizada por su contenido, dirección y fuerza.
	<i>Lazos</i>	Un lazo conecta a un par de actores a través de 1 o más relaciones. También se pueden describir en base a su contenido, dirección y fuerza.
	<i>Multiplicidad</i>	Entre más relaciones posee un lazo, entonces más multiplicidad hay en el lazo.
	<i>Composición</i>	La composición de una relación o lazo es derivada de los atributos sociales de ambos participantes.
Wellman [7]	<i>Rango</i>	Se refiere al tamaño y heterogeneidad de la red social.
	<i>Centralidad</i>	Se refiere a quién es central en la red social.
	<i>Roles</i>	Se refiere a la similitud en el comportamiento de los miembros de la red.
	<i>Grupos</i>	Es una estructura que se descubre en forma empírica, a través del análisis de los patrones de relaciones entre los miembros de la población. Los grupos emergen como conjuntos altamente conectados de actores.
<i>Comunidades Virtuales</i>		
Preece [44]	<i>Personas</i>	Corresponde a las personas que interactúan con otras en la comunidad, quienes poseen necesidades individuales, sociales y organizacionales.
	<i>Propósito</i>	Una comunidad comparte un interés, necesidad, información, servicio, o soporte, el cual provee un motivo a sus miembros para pertenecer a la comunidad.
	<i>Políticas</i>	El lenguaje y protocolo que guían las interacciones entre las personas y contribuyen al desarrollo del “folclore” o rituales que dan un sentido de normas sociales aceptadas e

		históricas.
Henri & Pudelko [4]	<i>Intención</i>	También denominado objetivos de la comunidad.
	<i>Inicio</i>	Los métodos a través de los cuales se crea el grupo inicial que conforma la comunidad.
	<i>Evolución</i>	Se refiere a la evolución temporal, tanto de la intención como de los métodos que forman grupos al interior de la comunidad.
Kim [3]	<i>Propósito</i>	Se refiere a los objetivos de los miembros de la comunidad.
	<i>Lugares</i>	Se refiere a los lugares en los cuales los miembros trabajan en conjunto.
	<i>Perfiles</i>	Es un conjunto de información que dice algo acerca de quién es un miembro, en el contexto de la comunidad.
	<i>Roles</i>	Se refiere a la existencia de distintos tipos de miembros al interior de la comunidad.
	<i>Liderazgo</i>	Se refiere al más visible de los roles. Corresponde a aquellos miembros que ayudan a mantener la comunidad funcionando.
	<i>Etiquette</i>	Es un conjunto de comportamientos – o estándares dentro de la comunidad – que un grupo de personas acepta.
	<i>Eventos</i>	Se refiere a las reuniones al interior de la comunidad, las cuales ayudan a definir la comunidad, recordarles a sus miembros lo que tienen en común y de qué se trata la comunidad en la cual participan.
	<i>Rituales</i>	Corresponden a celebraciones de carácter social (por ejemplo el cumpleaños de uno de los miembros), los cuales permiten crear un sentido de identidad al interior de la comunidad.
	<i>Subgrupos</i>	Se refiere a grupos pequeños de personas, en los cuales sus miembros forman sus más profundas relaciones y fuertes lealtades.
Selznik [47], Schwier [115]	<i>Historia</i>	La existencia de historias y cultura compartidas son un elemento que fortalece los lazos comunitarios.
	<i>Identidad</i>	Las comunidades promueven un sentido de identidad compartida.
	<i>Mutualidad</i>	Las comunidades brotan y se mantienen por las interdependencias y reciprocidades existentes.
	<i>Pluralidad</i>	Las comunidades obtienen gran parte de su vitalidad al ser “asociaciones intermedias” entre familia, iglesias y otros

		grupos periféricos.
	<i>Autonomía</i>	A pesar de haber un énfasis en la identidad grupal, es importante respetar y proteger la identidad individual de cada miembro.
	<i>Participación</i>	La participación social en la comunidad, especialmente aquella que promueve la autodeterminación, apoya la autonomía y permite que la comunidad se sostenga.
	<i>Integración</i>	Todos los otros elementos dependen de las normas de apoyo, las creencias y prácticas.
<i>Comunidades de Práctica (CoP)</i>		
Wenger [39]	<i>Comunidad</i>	Se refiere a las relaciones que edifican los miembros para poder aprender entre sí.
	<i>Práctica</i>	Se refiere a un repertorio compartido de recursos: experiencias, historias, herramientas, las maneras de abordar los problemas recurrentes.
	<i>Dominio</i>	Una CoP tiene una identidad definida por un dominio compartido de interés.
Plaskoff [42]	<i>Believing</i>	Se refiere a la idea de que los miembros necesitan creer en el valor intrínseco de la comunidad.
	<i>Behaving</i>	Indica que los miembros a desarrollar y seguir las normas de una comunidad.
	<i>Belonging</i>	Esto significa que los miembros de cultivar un sentimiento de pertenencia a una comunidad.
Lesser [48]	<i>Conexiones</i>	Existe una serie de conexiones entre los individuos de la comunidad. En otras palabras, los individuos perciben que ellos mismos son parte de una red (dimensión de estructura).
	<i>Relaciones</i>	Corresponde a un sentido de identidad que es desarrollado a través de las conexiones (dimensión de relaciones).
	<i>Contexto Común</i>	Los miembros tienen un interés común o un entendimiento compartido del ambiente en el que participan (su organización, por ejemplo) (dimensión cognitiva).
Saint-Onge [45]	<i>Práctica</i>	Se refiere a las actividades que realizan los miembros.
	<i>Personas</i>	Se refiere a quienes están involucrados en la comunidad.
	<i>Capacidades</i>	Se refiere a la capacidad de aprovechar las ventajas competitivas que poseen.

Anexo C. Caracterización de las Redes Sociales Generadas para la Comunidad I

	reply_all		creator		reply_prev		reply_all30	
	v	e	v	e	v	e	v	e
2009-01	110	1.517	106	278	104	398	105	803
2009-02	92	1.238	90	210	92	312	92	959
2009-03	99	1.518	96	280	99	417	97	1.166
2009-04	90	1.066	86	233	89	372	89	786
2009-05	99	1.277	89	198	94	346	94	824
2009-06	94	1.103	90	229	94	379	92	909
2009-07	94	1.053	87	223	90	341	88	809
2009-08	94	938	92	212	92	323	92	694
2009-09	79	717	76	181	79	294	79	621
2009-10	86	821	84	172	86	267	85	634
2009-11	92	773	87	185	91	297	90	663
2009-12	85	778	78	173	83	301	84	598
2010-01	77	692	75	188	77	273	75	577
2010-02	72	770	70	158	72	245	72	639
2010-03	82	592	74	121	80	205	79	473
2010-04	82	619	76	181	81	296	80	520
2010-05	88	761	83	178	87	288	87	583
2010-06	81	1.217	78	239	79	366	77	928
2010-07	103	1.470	98	319	102	492	101	1.305
2010-08	101	934	94	195	100	316	96	700
2010-09	74	979	69	164	72	246	72	797
2010-10	87	660	80	147	83	252	83	520
2010-11	73	694	69	165	70	255	69	577
2010-12	86	865	78	178	83	285	82	623

Anexo D. Tiempos de Procesamiento para la técnica PYTM para la Comunidad I

n_topicos	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	TOTAL
5	116	502	752	1.154	1.396	2.077	1.538	876	930	788	469	288	132	49	11.074
10	175	724	1.067	1.663	1.930	2.882	2.183	1.214	1.293	1.069	663	420	196	69	15.561
15	244	960	1.408	2.161	2.539	3.686	2.870	1.540	1.780	1.493	903	560	264	93	20.520
20	310	1.257	1.784	2.770	3.143	4.573	3.575	2.013	2.202	1.856	1.164	740	346	117	25.875
25	373	1.565	2.266	3.338	3.902	5.471	4.312	2.465	2.731	2.329	1.440	906	417	142	31.688
30	456	1.933	2.853	4.000	4.623	6.763	5.122	3.056	3.211	2.784	1.691	1.054	490	164	38.238
35	518	2.278	3.315	4.776	5.428	7.862	5.923	3.416	3.853	3.209	1.912	1.192	552	185	44.463
40	591	2.591	3.792	5.441	6.170	8.867	6.786	4.005	4.248	3.681	2.223	1.391	639	215	50.690
45	655	2.944	4.202	6.091	7.189	10.063	7.655	4.455	4.941	4.094	2.443	1.525	696	231	57.240
50	724	3.249	4.595	6.871	7.838	11.290	8.494	4.990	5.308	4.576	2.755	1.723	782	257	63.514
55	780	3.546	5.018	7.560	8.618	12.280	9.476	5.396	6.035	5.059	2.977	1.837	841	278	69.769
60	859	3.867	5.516	8.270	9.402	13.552	10.381	5.981	6.503	5.530	3.300	2.050	930	304	76.519
65	922	4.181	5.903	8.924	10.092	14.372	11.120	6.411	7.130	5.902	3.515	2.172	983	325	82.032
70	994	4.461	6.357	9.564	11.002	15.662	12.134	6.873	7.495	6.504	3.919	2.355	1.066	351	88.824
75	1.085	4.778	6.883	10.310	11.722	16.753	12.581	7.471	8.211	6.819	4.053	2.468	1.114	374	94.715
80	1.137	5.136	7.329	11.147	12.553	18.023	13.690	7.846	8.770	7.312	4.403	2.730	1.220	401	101.796
85	1.209	5.385	7.740	11.657	13.147	18.924	14.386	8.623	9.468	7.605	4.539	2.795	1.280	423	107.286
90	1.278	5.695	8.100	12.363	14.065	20.891	15.586	8.857	9.675	8.175	4.847	3.021	1.355	451	114.470
95	1.351	6.054	8.823	13.079	14.839	21.554	15.964	9.520	10.543	8.471	5.113	3.164	1.426	464	120.483
100	1.411	6.395	9.218	13.838	15.656	24.049	17.368	9.979	10.728	9.077	5.556	3.409	1.515	497	128.820
	4,22	18,75	26,92	40,27	45,90	66,55	50,32	29,16	31,96	26,76	16,08	9,94	4,51	1,50	372,85