Bibliografía Anotada

Juan David Cruz Gómez, 299618
Maestría en Ingeniería – Ingeniería de Sistemas y Computación
Seminario de Investigación I

27 de Septiembre de 2005

1. Privacy-Preserving Data Mining [2]

Este artículo trata sobre como preservar la privacidad de usuarios durante un proceso de minería de datos. Lo que se propone es utilizar una perturbación generada por una distribución gaussiana o uniforme sobre aquellos datos que las personas quieren que permanezcan ocultos. El problema que plantean, es que si un número muy alto de usuarios realiza esta perturbación, ¿es posible reconstruir los datos con suficiente exactitud?.

Para esto plantean dos formas de realizar la clasificación de los datos, el problema es que solo funciona, por ahora, con datos numéricos.

Puede ser interesante, podría incluirse dentro de alguna categoría de técnicas de reconstrucción o limpieza de datos durante la creación del estado del arte en la parte de minería de datos.

2. Data miming for hypertext: A tutorial survey [13]

Este es un "survey" sobre diferentes métodos para extraer conocimiento de textos, específicamente de hipertextos. El autor divide la discusión en cinco partes:

- Modelos básicos: en donde explica algunos modelos para representar y analizar textos, hipertextos y datos semiestructurados. La idea de esots modelos es encontrar una representación de todo un documento, de modo que se pueda extraer conocimiento relevante de ellos.

  - Modelos para texto: El primer enfoque para este modelo es el de tomar cada una de las palabras de un documento y convertir cada una de ellas en un eje de un espacio euclidiano. Así, si en un documento d, una palabra t ocurre $n(d,t)$ veces, ese será el valor en el eje que representa dicha palabra, en decir, se convierte el documento en un vector $m-$dimensional, en donde m es el número de diferentes palabras que contenga. Con esta representación vectorial es natural pensar en realizar una normalización, para lo que se utiliza alguna de las siguientes normas:

$$\|d\|_1 = \sum_t n(d,t)$$

$$\|d\|_2 = \sqrt{\sum_t n(d,t)^2}$$

$$\|d\|_\infty = \max_t n(d,t)$$

El problema con esta representación es que no tiene en cuenta la importancia de las palabras sino que le asigna el mismo peso a todas. Entonces, los ejes se modifican estirándolos, dependiendo de cada palabra, para eso primero se busca una medida de la rareza que tiene una palabra, esto está dado por:

$$r = \frac{N_t}{N}$$

donde N es el número de documentos que contienen la palabra t. Luego se utiliza una función llamada *frecuencia inversa del documento*, la cual está dada por:

$$IDF(t) = 1 + \log(r)$$

Se pueden construir también modelos probabilísticos, basados en la probabilidad θ_t de que al lanzar un dado aparezca la palabra t. Me parece que la mejor forma de representar estos modelos probabilísticos es utilizando un modelo oculto de Markov.
El principal problema de los modelos para texto es que no capturan la semándica de los documentos, así que no importa nunca el orden las palabras, esto es llamado *una bolsa de palabras*.

- Modelos para hipertexto: los hipertextos, además de contener texto, tienen hipervínculos, lo que hace que estén relacionados con otros documentos en algún contexto determinado. Estas relaciones se pueden representar fácilmente utilizando un grafo dirigido (D, L), en donde D es el conjunto de nodos, en este caso, documentos y L es el conjunto de vínculos entre ellos.
  Este grafo se puede ver como una red social, aunque en este caso, no se una red de personas sino de documentos, pero al igual que la de personas, es creada en torno a un tema específico y bien definido (aunque se pueden presentar casos en que esto no sea cierto).
- Modelos para datos semi-estructurados: los diferentes documentos pueden tener vínculos tanto dentro del documento como hacía otros, se tienen entonces directorios Web, que crean una relación de la forma *es_un(tópico_específico, tópico_general)*, y un *ejemplo(tópico, URL)*, en dónde una URL es tratado como un tópico más. Los datos semi-estructurados seon el punto de convergencia entre la Web y las comunidades de bases de datos, los primeros trabajan con documentos y los segundos con datos. Dichos datos están evolucionando rápidamente hacia formas más flexibles que el modelo relacional, como XML.

- Aprendizaje supervisado: se explican algunas técnicas de clasificación, en donde, en general, el entrenamiento se realiza con datos maracdos de alguna forma.

  - Modelos Probabilísticos: sean m clases o tópicos $c_1, c_2, ..., c_m$, con algún documento asociado a cada clase D_c, la probabilidad de una clase determinada está dada por: $|D_c| / \sum_c |D_c|$. Se define T como el universo de palabras, vocabulario.
    - Clasificación de Bayes *pura:* Los modelos de bolsa de palabras utilizan de forma natural esta forma de clasificación (En inglés, se denomina *naive Bayes classification*, y naive se refieres a que asume la independencia de cada uno de los términos). hay algunos desarrollos matemáticos importantes, se deben revisar, de modo que se puedan incluir en un estado del arte.

    - Selección de Características y suavizado de parámetros: es una técnica que incluye la idea de *stopwords*, que son aquellas palabras que no se tendrán en cuenta para calcular la probabilidad. Hay otras consideraciones para calcular dicha probabilidad, es bueno revisar estas fórmulas, la que no son muy claras.
    - Modelo de dependencia limitada: En este modelo plantean la utilización de una red de bayes. No es muy claro en esta red que es un nodo y como están conectados cada uno de ellos.
    - Otros modelos avanzados: Hablan, pero no profundizan en técnicas como la de máxima entropía, support vector machines, árboles de desición, perceptrones, en general, aquelas técnicas que buscan la forma de separar un espacio y las que no lo hacen
  - Relaciones de aprendizaje: con esto se buscan las relaciones entre documentos a partir de los enlaces que cada uno de ellos tiene. Es interesante, pero lo explican muy supercialmente, tocaría revisar la bibliografía.

- Aprendizaje no supervisado: en donde explican algunas técinicas de agrupación, la idea es que un clasificador de este estilo tome hipertextos y genere una jerarquía que permita agrupar los documentos según sus características.

  - Técnicas básicas de agrupamiento: se asume que los documentos son representados como vectores, de modo que se pueden trabajar en un espcio $m-$dimensional, donde m corresponde al número de palabras en el documento.
    - $k-$means: en esta técnica, se asume que se requieren k grupos, para esto se toman k *documentos semilla*, los cuales son elegidos arbitrariamente. De forma iterativa, cada documento que se tiene como entrada, es asignado al documento semilla más similar. Las coodenadas de la semilla es recalculada de modo que sea el centroide de todos los documentos asignados a ella; este proceso es repetido hasta que se estabilizan las coordenadas de la semilla. El problema de esta ténica es que es dependiente del número de palabras en cada documentos, es decir, el orden de ejecución es dependiente de la dimensionalidad.
    - Agrupación aglomerativa: esta técnica fusiona documentos en super-documentos o grupos, lo que genear una jerarquía basada en la similaridad. La similaridad es definida por el coseno del ángulo de los vectores definidos por cada documento en el espacio vectorial.
  - Técnicas basadas en álgebra lineal: debido a la representación de documentos como vectores, es posible aplicar transformaciones a cada documento utilizando técnicas derivadas del álgebra lineal.
    - Indexado semántico latente: la similitud entre dos documentos vistos como vectores se refiere principalmenet a la similitud sintáctica, pero no tiene en cuenta la similitud semántica, es decir, tiene en cuenta el como, pero no su significado, por ejemplo, cuando dos documentos tratan de un mismo tema con palabras diferentes, (carro, automóvil, transmisión, caja de velocidades).
    - Proyecciones aleatorias: esta técnica busca reducir el tiempo de cálculo de la similitud de documentos, lo que hace es reducir la dimensionalidadde cada punto haciaendo algunas transformaciones sobre un espacio generado de forma aleatoria.

- Aprendizaje Semi-supervisado: esta técnicas buscan cubrir las posibles fallas y debilidades de las técnias que utilizan aprendizaje supervisado y no supervisado.

    - Aprendizaje a partir de documentos etiquetados y no etiquetados: la idea es utilizar primero la técnica de clasificación bayesiana pura, esto es, para los datos etiquetados, y luego utilizar una técnica de agrupación.

- Análisis de redes sociales: el análisis de redes es una teoría que se basa en la conectividad y distancias entre nodos de un grafo. En este caso es aplicada a páginas Web.

    - Google: explican la forma en que Google realiza la búsquedas basado en la medida *page rank* y explican porque las búsquedas se realizan tan rápido.
    - Búsqueda de hipervínculos inducida por tópicos: esta técnica es utilizada por otros buscadores en Internet, los cuales no están buscando de forma constante sino que en el momento de recibir una consulta, toman un sub-grafo y sobre este realizan la búsqueda.

Este artículo es bueno, es un poco corto para ser un survey, sin embargo muestra las técnicas más utilizadas para realizar búsquedas de y entre documentos, es importante. No tiene en cuanta cosas como la maldición de la dimensionalidad.

3. Knowledge Management: A new Idea or a Recycled Concept? [38]

Este no es un artículo, es más que eso. Es un estudio sobre la gestión del conocimiento, cuales spn sus fases, que estudios se han hecho sobre eso. Tiene diferentes puntos de vista: datos e información, capital intelectual y mapas de conocimiento, tecnología y su influencia sobre el conocimiento, modelos y el proceso de transformación del conocimiento. Es interesante, ya que no es un enfoque desde el punto de vista de ciencias de la computación, sino desde la administración y las ciencias de la información.

Es definitivamente un documento clave para armar la parte sobre gestión del conocimiento del marco teórico.

4. Researching Organizational Systems using Social Network Analysis [44]

Este artículo intenta analizar las organizaciones utilizando como herramienta el análisis de redes sociales, la diferencia es que lo hacer desde el punto de vista de los sistemas de información y añade además todos aquellos aspectos de las organizaciones sociales reales, es decir, todo aquello que influye en la vida real en la comunicación itra-organizacional y que se puede ver la utilización que se le da a los diferentes medios electrónicos de comunicación.

Aunuq eno tiene un componente matemático fuerte, de hecho, no tiene, es un punto de vista interesante de las organizaciones y como se pueden crear redes a partir de sus sistemas de inforamción, además aporta teoría interesante sobre el análisis de redes sociales, puede apoyar mucho la creación del estado del arte en la parte de ARS y su utilización.

5. Visual mapping of articulable tacit knowledge [11]

Este artículo cuenta de forma clara que es el conocimiento tácito o suave, y realiza un gran énfasis en el hecho de que la gestión del conocimiento que se hace en el mundo occidental es muy diferentes

al que se realiza en el mundo oriental. Hay una bibliografía extensa, en domde nombra autores muy reconocidos en materia de gestión de conocimiento, como Nonaka y Takeuchi.

Algo muy interesante que hacen es crear una tabla con todo aquello que puede ser tratado como conocimiento tácito, pero además, la dividen en el conocimiento tácito que puede ser articulado y el que no.

Luego proponen como herramienta para gestionar el conocimiento, el análisis de redes sociales, de donde obtienen información sobre el flujo de conocimiento en una organización y como puede estar este entorno a una persona o *portero*, como lo llaman ellos.

Es muy interesante, creo que realiza aportes muy interesantes en la teoría sobre gestión de conocimiento y muestra un caso de estudio general para el análisis de redes sociales.

6. Organization of Social Knowledge in Multi-agent Systems [33]

Este artículo muestra como utilizar el conocimiento social para diseñar sistemas multi-agentes. Es interesante ya que muestra como utilizar y organizar todo ese conocimiento generado en una organización X con el fin de generar un modelo válido de un grupo de agentes que interactúan.

El problema general de estos métodos es que no solo traen lo bueno de las organizaciones socieles, sino que también traen lo malo, como por ejemplo los cuellos de botella y cosas por el estilo.

Es interesante, pero por ahora no es muy útil.

7. Mining the network value of custumers [17]

Este artículo trata sobre minería de datos para obtener conocimiento de la red de valor de los clientes de un negocio determinado. Se basa en el hecho de que los clientes se pueden ver como una red social y que en dicha red el conocimiento se difunde como una epidemia, como lo haría un virus. Cada nodo en la red es un cliente y la red es vista como un campo aleatorio de Markov, y bajo este hecho desarrollan toda su discusión.

Aunque es interesante, no me parece muy significativo en el campo de ARS, depronto podría servir en la parte de minería, como una técnica muy específica para obtener conocimiento de datos muy específicos.

8. Visualization of Bibliographic Networks with a Reshaped Landscape Metaphor [9]

Es una rticulo muy interesante, que pone una red de citaciones de artículos científicos en un paisaje, de modo que se genera un dibujo motañoso tridimensional en dónde la altura le da la importancia a un artículo determinado.

Aunque es un método muy interesante de agrupación para este tipo de aplicación, es computacionalmete pesado, no solo en el tratamiento de los datos, sino en la generación del paisaje. Además creo, que analizar un paisaje es mucho más complicado que analizar un conjunto de datos organizados en tablas.

Puede incluirse dentro de las técnicas de agrupamiento, como una cita o algo así, pero no es un método que sea fácilmente aplicable a cualquier problema.

9. The integration of business intelligence and knowledge management [15]

Este artículo habla sobre las dos tendencias de manejo de información: business intelligence y knowledge management, muestra como se pueden combinar para extraer los mayores beneficios. Es bastante interesante ya que muestra el punto de vista empresarial y como se le puede dar un mayor valor a la información disponible tanto dentro de la compañía como fuera de ella. Puede aportar bastante en el marco teórico, en la parte de gestión de conocimiento.

10. The Structure of Broad Topics on the Web [14]

Este artículo presenta la estructura de Internet como un gran grafo y discute sobre algunas técnicas utilizadas para recorerlo. Aunque es relativamente reciente, no habla de Google, aunque si habla del PageRank, también habla de Yahoo y de Open Directory Project.

Es bastante denso, plantea una distribución de tópicos, según la cual se propone una nueva forma de recorer el grafo.

Podría servir para el marco teórico, pero es muy específico, plantea muchos desarrollos matemáticos que no son simples.

11. Activity theory and distributed cognition: or What does CSCW need to do with theories? [22]

Es un ensayo, muy denso, muy largo, hay que leerlo con mucha calma, puede ofrecer algunos componentes teóricos interesantes.

12. The Dynamics of Cultural Influence Networks [24]

Es un poster, pero nombre otro artículo en donde se hace una discusión más grande acerca de la influencia en una red social que hace que el conocimiento fluya a través de ella y sea una especie de conocimiento distribuído.

13. Efficient Management of Persistent Knowledge [27]

En este articulo proponen una nueva forma de realizar indexamiento de grandes volúmenes de datos en fuentes para motores deductivos llamada árboles G, de modo que la información obtenida de una consulta sea lo suficientemente buena para el que la realizó.

No es muy relevante dado que simplemente explica como almacenar y extraer información de una estructura de datos pero no muestra nada sobre como utilizar el conocimiento.

14. An ontology model to facilitate knowledge sharing in multi-agent systems [41]

Este artículo presenta un modelo de conocmiento basado en las ontologías de documentos y de comunicaciones y lo aplican para modelar las interacciones de sistemas multi-agentes. Es un artículo muy interesante, ya que propone una forma diferente para modelar este tipo de interacciones, pero no aporta mucho para el trabajo, o al menos para el marco teórico no.

15. Mining Newsgroups Using Networks Arising From Social Behavior [1]

Este artículo muestra una técnica para extraer conocimiento de las redes sociales que se forman en grupos de noticias. Anuque se eligen los grupos de noticias, el tratamiento final se realiza sobre una red social, así que las técnicas en principio son válidas también para cualquier red , sin importar la forma en la que se genere. Puede aportar algo de teoría sobre grafos, pero no teoría tratada normalmente, sino vista desde las redes sociales.

16. Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study [3]

Este artículo realiza un interesante estudio empírico sobre las redes sociales presentes en una organización. Las redes son generadas a partir de los correos electrónicos y utlizan como medida de la red la centralidad. Esta medida se refiere a al grado de perteneneciia de un individuo a un grupo, o a un grupo virtual, en este caso. Además utiliza otros conceptos organizacionales: *rol funcional, estatus (dentro del grupo virtual), rol de comunicación*, los que también son incluídos en el calculo del desempeño individual.

Es un artículo muy interesante ya que no solo trata de forma teórica el análisis de redes sociales sino que aplica su estudio a un caso conreto en donde miden el desempeño de un inviduo en una red. Puede aportar mucho en el marco teórico, desde el punto de vista de las redes sociales y en técnicas de análisis.

17. The Design of Discovery Net: Towards Open Grid Services for Knowledge Discovery [4]

Este artículo presenta el diseño y un caso de estudio sobre el almacenamiento y recuperación de conocimiento de datos distribuidos en una red. Todo el estudio está hecho en torno a las proteínas y la iformación que pueda proveer el ADN.

No es muy interesante para el marco teórico, sin embargo, podría ser utilizado para descubrir conocimiento en una red social (ambiente distribuído).

18. Experiments in Social Data Mining: The TopicShop System [5]

Este artículo es muy interesante, explica un sistema para extraer información social. Busca responder dos preguntas: (1) ¿Es la información extraida valiosa? y (2) ¿Las interfaces basadas en iformación mejoran la productividad de los usuarios?. Lo que proponene es un sistema de búsqueda de información basado en los tópicos de las páginas Web. Es un concepto interesante, ya que lo que hace primnero es una agrupación de dichos tópicos, que es algo que se debe realizar cuando se vaya a hacer el análisis de los correos para construir las redes sociales. Es para tener en cuenta en la parte de diseño de la herramienta propuesta, se pueden sacar ideas muy interesantes para la agrupación.

19. A Relational View of Information Seeking and Learning in Social Networks [6]

Es un artículo muy interesante, habla de redes sociales, pero no de análisis de redes sociales, o al menos de la teoría que rodea esto. Sin embargo aborda el tema de la obtención de información desde el punto de vista subjetivo de las relaciones reales entre personas. Además hace énfasis en los lazos fuertes y débiles y como influyen en una red social para que la comunicación y la información sea compartida de forma exitosa.

Es teóricamente interesante, puede aportar algunas ideas sobre análisis en el marco teórico.

20. Visual Unrolling of Network Evolution and the Analysis of Dynamic Discourse [8]

Este es un artículo que expone una técnica bastante interesante para evaluar discursos, pero no solo discursos aislados, sino aquellos que están en una red, por ejemplo una lista de discusión o un grupo de noticias. Es bastante interesante para la fase de construcción del prototipo ya que se podría utilizar para agrupar conversaciones o para realizar un filtrado inicial. Está diseñado de tal forma que el método sea aplicado a cualquier tipo de red (grafo).

21. The graphical interpretation of plausible tacit knowledge flows [10]

En este artículo crean redes sociales de conocimiento pero utilizando el flujo de conocimiento tácito. Esto no es una tarea simple; para capturar dicho conocimiento diseñaron unos formularios con ciertas preguntas muy específicas. Personalmente creo que formularios no es la forma de extraer conocimiento tácito, sin embargo, el trabajo es muy interesante ya que hace una división muy clara entre los tipos de conocimieto y además añade al conocimiento explícito una división: el conocimiento codificado, que es todo aquel que está escrito de alguna forma. Más adelante utiliza el análisis de redes sociales para analizar las redes de conocimiento tácito que han generado, tienen resultados interesantes.

22. Expertise Identification using Email Communications [12]

Este artículo trata sobre la identificación de expertos en un tema a través de los correos electrónicos. La pregunta que plantea para resolver es ¿quién sabe qué?, y para esto utiliza dos métodos de identificación: el primero es analizar los textos de los correos y el segundo es analizar el grafo generado por la red social. Luego a ambos métodos le aplica una técnica de detección de señales llamada (en el paper) d', con la que se sabe que tan acertada fué la técnica. Finalmente concluyen que la mejor identificación se logra cuando se analiza el grafo.

Es muy interesante, aporta mucho, lo malo es que es un poco corto y en algunos casos la profundidad no es mucha.

23. Building Knowledge Discovery into a Geospatial Decision Support System [23]

Este paper no me pareció interesante, explica una técnica de minería de datos para generar reglas de sociación de dots geoespaciales. En realidad no me parece que aporte mucho al trabajo.

24. Maximizing the Spread of Influence through a Social Network [28]

Este artículo trata sobre la difusión de información en diferentes ámbitos de estudio y como esta difusión influencia a las personas en dichos ámbitos. Lo que proponen es un método para maximizar la influencia de la difusión de información en una red social específica, para esto, primero realizan una definición algo rigurosa de los términos que rodean el ARS y luego proponen su método con bastante rigurosidad matemática. Es muy interesante y puede aportar al marco teórico, además que no es un artículo de solo "carreta", sino que presenta formalmente los conceptos y desarrollos.

25. The Link Prediction Problem for Social Networks [31]

Este articulo comienza planteandose la siguiente pregunta: "Dada una foto de una red social en un tiempo t, ¿se puede predecir acertadamente la estructura de la red en un tiempo $t' = t + 1$?", es decir, ¿se podrá inferir que nuevas relaciones entre sus miebros pueden aparecer?. Esto no es sencillo, se plantea aquí que si se puede y se proponen algunos métodos para hacerlo. El problema es que aunque muestran algunos resultados positivos, no tienen conclusiones, hubiera sido interesante que las tuviera. Sin embargo, es muy bueno, puede paortar cosas interesantes durante el procesos de desarrollo de la herramienta en una fase posterior del proyecto.

26. Recommending Collaboration with Social Networks: A Comparative Evaluation [34]

Este artículo está enfocado al trabajo colaborativo asistido por computador (CSCW por sus iniciales en inglés). Y muetra como recomendar colaboraciones utilizando las redes sociales y su análisis. Es un artículo que no tienen mucho fundamento matemático (no quiere decir que sea malo) ya que está

visto desde el concepto de organización y los grupos que se pueden forma allí y no desde el análisis matemáticos de grafos. Es teóricamente fuerte, por lo que puede ayudar en la construcción del estado del arte tanto para larte de redes sociales como para la parte de trabajo colaborativo.

27. Studying Recommendation Algorithms by Graph Analysis [35]

Este artículo describe una forma de recomendación aplicada a comercio (recomendación personalizada) a partir de la información proporcionada por una red social. Este análisis es realizado diréctamente sobre el grafo, es decir, lo hace de forma matemática, analizando los saltos entre personas. Muestra conceptos interesantes, pero no aporta mucho para el marco teórico.

28. AISEC: an Artificial Immune System for E-mail Classification [37]

Este artículo presenta una forma de clasificación automática de correos utilizan sistemas inmunológicos artificiales. Es muy interesante. Lo que se propone es un esquema de selección clonal, en donde se utiliza el esquema de entrenamiento utilizado comunmente. Los correos nuevos son tratados como antígenos, que luego son presentados a todo el sistema inmunológico. Algo muy interesante es la representación de las células, en donde solo se tienen en cuenta las palabras que van en el remitente y en el asunto, logrando que el número de células no sea muy alto (algo interesante para un AIS). Hay que estudiarlo con más profundidad y revisar otros por el estilo.

29. Technology and knowledge: bridging a generating gap [39]

Este artículo presentala brecha existente entre tecnología e información, sus diferenecias y como se podría reducir dicha brecha. Es tratado desde un punto de vista comercial, se habla de CRM y de técnicas de minería de datos. Aporta información interesante sobre gestión de conocimiento, tecnología y sus relaciones en la empresa, pero no aporta mucho en análisis de redes sociales.

30. Archipelago: A Network Security Analysis Tool [40]

Este artículo expone un sistema para realizar análisis de seguridad en redes, pero no solo en redes de computadores, en general, a cualquier tipo de grafo, pero haciendolo como si estuviera analizando redes sociales. Lo interesante es que muestra muchas medidas de redes sociales, como el grado o la centralidad de un nodo.

Puede aportar mucho a la teoría de redes sociales, análisis de grafos y a aplicaciones.

31. Searching Social Networks [43]

Este artículo muestra otra técnica para realizar búsquedas en redes sociales, específicamente en redes de referencias. Es interesante, ya que muestra diferentes aproximaciones al problema y las compara, lo

que puede ser muy útil en el marco teórico y en el momento de evealuación de técnicas para extraer información y de identificar personas dentro de la red.

32. Extracting social networks and contact information from email and the Web [16]

Este artículo muestra un sistema de recuperación de información a partir de los correos electrónicos. Lo que hace es tomar a cada una de las personas en un mensaje de correo y obtener la información de contacto de la Web. Plantean un modelo probabilístico muy interesante para hacerlo. Es muy interesante para el marco teórico, tanto para ARS como para aplicaciones.

33. Structural and Algorithmic Aspects of Massive Social Networks [18]

Este artículo muestra un estudio sobre grandes redes sociales, está muy enfocado al estudio matemático de este tipo de estructuras. No aporta mucho a lo que ya se sabe sobre el análisis de redes sociales, solo entrega un modelo matemático para generar grafos muy grandes y simular su comportamiento.

34. Reputation Network Analysis for Email Filtering [19]

Este artículo muestra otra forma para realizar la clasificicación de correos electrónico, pero basada en la reputación de las personas de la red. Esta reputación es econtrada a partir de las relaciones existentes en la red, no solo las de un grado, sino más, lo que permite establecer si una persona con la que no se está conectada directamente es o no confiable. Para demostrarlo, desarrollaron un prototipo cliente de correo, el cual está en capacidad de encontrar la confiabilidad de un mensaje.

Es muy interesante, no solo provee información para el marco teórico sino que aporta información sobre la aplicaciones que tiene.

35. Multiple Email Addresses: A Socio-technical Investigation [21]

Este artículo presenta un estudio socio-técnico acerca de la utilización de difrentes direcciones de correo electrónico por una sola persona. No me parece que aporte mucho, o al menos en esta etapa del proyecto no. No vale la pena desecharlo, puede dar algunas ideas de comportamiento que podrían ayudar al diseño de la herramienta en una etapa posterior.

36. Enhancing Reputation Mechanisms via Online Social Networks [25]

Este artículo (es un poster) muestra la forma de utilizar redes sociales para asegurar la confiabilidad de un individuo o de un grupo. Propone dos técnicas: PageRak y una basada en los pesos de las

conexiones de la red social en al que aparece dicho individuo. No aporta mucho, en realidad solo son ideas, habría que buscar el paper completo, de otro modo no sirve.

37. Six Degrees of Jonathan Grudin: A Social Network Analysis of the Evolution and Impact of CSCW Research [26]

Este artículo muestra el funcionamiento de redes sociales en grupos de trabajo colaborativo. La idea es evaluar el comportamiento de las redes de co-autoría de artículos. No proponene ningún modelo matemático para realizar el análisis, sin embargo utilizan una muy bien fundamentada teoría sobre el ARS. Puede aportar mucho al marco teórico, además hay que tener en cuenta que es relativamente reciente.

38. Editorial: Data Mining Lessons Learned [30]

Este editorial realiza un estudio superficial de diferentes técnicas de minería de datos, así como algunas cosas que han sido aprendidas durante el tiempo. Es interesante para delimitar el marco teórico en el campo de minería de datos. Debe ser tenido en cuenta durante la escritura de este.

39. Introduction to the CMOT Special Issue on Mathematical Representations and Models for the Analysis of Social Networks within and between Organizations [32]

Esta es la introducción de una publicación especial sobre representaciones matemáticas para el análisis de redes sociales. Lo interesante es que hace una revisión bibliográfica de los diferentes modelos, enfoques y teorías, lo que hace que sea un insumo importante para la construccción del estado del arte.

40. Dynamic Supernetworks for the Integration of Social Networks and Supply Chains with electronic Commerce: Modeling and Analysis of Buyer-Seller Relationships with Computations [42]

Este es un artículo bastante denso, plantea un modelo matemático de redes dinámicas para integrar redes sociales y redes de aprovisionamiento. No aporta mucho en conceptos y es muy específico como para tener en cuenta para el marco teórico. Podría ser utilizado como referencia de alguna aplicación o algo así.

41. PEEP – An Information Extraction based approach for Privacy Protection in Email [7]

Este artículo muestra una herramienta para extraer información de los correos salientes de una compañía y verificar si violan alguna política de seguridad y de manejo de información. Es bastante

completo, muestran la arquitectura del sistema y explican de forma detallada como extraen y verifican la información de los correos. Aunque es una aplicación más enfocada a correos, puede servir como ejemplo de formas de extracción de conocimiento de correos electrónicos.

42. Implicit Queries for Email [20]

Este artículo plantea la forma en que se pueden extraer frases o palabras relevantes del cuerpo de un correo electrónico para ser enviadas a un motor de búsqueda. Definen la forma de encontrar dichas cadenas de caracteres y también definen la forma en que dichas consultas van a ser construídas. Es interesante desde el punto de vista de extracción de información de los correos, lo otro es una utilidad adicional, que por el momento no es interesante.

43. Email Task Management: An Iterative Relational Learning Approach [29]

Este artículo plantea la forma de presentar los mensajes de correo electrónico de modo que no se interactúe con cada uno al tiempo sino que se pueda tener un mapa conceptual o una agrupación que permita hacerlo, es decir, agrupar los emnsajes según su intención o la tarea que cumplen para el destinatario.

Es otra herramienta que permite una ágil administración de los correos para reducir el tiempo que puede gastar un usuario qaue recibe un volumen alto de mensajes diarios.

44. The Social Network and Relationship Finder: Social Sorting for Email Triage [36]

Este artículo plantea una idea interesante sobre el manejo automatizado de correos electrónicos, lo malo es que solo se limitan a mostrar la herramienta y los resultados obtenidos, no explican ninguna técnica ni nada por el estilo. Es útil en la medida que se utilice como punto de partida para generar nuevas aplicaciones o mejoras que se puedan desarrollar a lo que plantean.

Referencias

[1] Agrawal, R., Rajagopalan, S., Srikant, R., and Xu, Y. Mining newsgroups using networks arising from social behavior. In *WWW '03: Proceedings of the 12th international conference on World Wide Web* (New York, NY, USA, 2003), ACM Press, pp. 529–535.

[2] Agrawal, R., and Srikant, R. Privacy-preserving data mining. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (New York, NY, USA, 2000), ACM Press, pp. 439–450.

[3] Ahuja, M. K., Galletta, D. F., and Carley, K. M. Individual centrality and performance in virtual r&d groups: An empirical study. *Manage. Sci. 49*, 1 (2003), 21–38.

[4] Alsairafi, S., Emmanouil, F.-S., Ghanem, M., Giannadakis, N., Guo, Y., Kalaitzopoulos, D., Osmond, M., Rowe, A., Syed, J., and Wendel, P. The design of discovery

net: Towards open grid services for knowledge discovery. *Int. J. High Perform. Comput. Appl. 17*, 3 (2003), 297–315.

[5] Amento, B., Terveen, L., Hill, W., Hix, D., and Schulman, R. Experiments in social data mining: The topicshop system. *ACM Trans. Comput.-Hum. Interact. 10*, 1 (2003), 54–85.

[6] Borgatti, S. P., and Cross, R. A relational view of information seeking and learning in social networks. *Manage. Sci. 49*, 4 (2003), 432–445.

[7] Boufaden, N., Elazmeh, W., Matwin, S., Ma, Y., El-Kadri, N., and Japkowicz, N. Peep – an information extraction based approach for privacy protection in email. In *Proceedings of Second Conference on Email and Anti-Spam CEAS 2005* (Stanford University, July 21 & 22 2005).

[8] Brandes, U., and Corman, S. R. Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization 2*, 1 (2003), 40–50.

[9] Brandes, U., and Willhalm, T. Visualization of bibliographic networks with a reshaped landscape metaphor. In *VISSYM '02: Proceedings of the symposium on Data Visualisation 2002* (Aire-la-Ville, Switzerland, Switzerland, 2002), Eurographics Association, pp. 159–ff.

[10] Busch, P., Richards, D., and Dampney, C. N. G. K. The graphical interpretation of plausible tacit knowledge flows. In *CRPITS '24: Proceedings of the Australian symposium on Information visualisation* (Darlinghurst, Australia, Australia, 2003), Australian Computer Society, Inc., pp. 37–46.

[11] Busch, P. A., Richards, D., and Dampney, C. N. G. K. Visual mapping of articulable tacit knowledge. In *CRPITS '01: Australian symposium on Information visualisation* (Darlinghurst, Australia, Australia, 2001), Australian Computer Society, Inc., pp. 37–47.

[12] Campbell, C. S., Maglio, P. P., Cozzi, A., and Dom, B. Expertise identification using email communications. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management* (New York, NY, USA, 2003), ACM Press, pp. 528–531.

[13] Chakrabarti, S. Data mining for hypertext: a tutorial survey. *SIGKDD Explor. Newsl. 1*, 2 (2000), 1–11.

[14] Chakrabarti, S., Joshi, M. M., Punera, K., and Pennock, D. M. The structure of broad topics on the web. In *WWW '02: Proceedings of the 11th international conference on World Wide Web* (New York, NY, USA, 2002), ACM Press, pp. 251–262.

[15] Cody, W. F., Kreulen, J. T., Krishna, V., and Spangler, W. S. The integration of business intelligence and knowledge management. *IBM Syst. J. 41*, 4 (2002), 697–713.

[16] Culotta, A., Bekkerman, R., and McCallum, A. Extracting social networks and contact information from email and the web. In *Proceedings of First Conference on Email and Anti-Spam (CEAS) 2004* (Mountain View, CA, July 2004).

[17] Domingos, P., and Richardson, M. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2001), ACM Press, pp. 57–66.

[18] Eubank, S., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., and Wang, N. Structural and algorithmic aspects of massive social networks. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms* (Philadelphia, PA, USA, 2004), Society for Industrial and Applied Mathematics, pp. 718–727.

[19] Golbeck, J., and Hendler, J. Reputation network analysis for email filtering. In *Proceedings of First Conference on Email and Anti-Spam (CEAS)* (Mountain View, CA, July 30 – 31 2004).

[20] Goodman, J., and Carvalho, V. R. Implicit queries for email. In *Proceedings of Second Conference on Email and Anti-Spam CEAS 2005* (Stanford University, July 21 & 22 2005).

[21] Gross, B. M. Multiple email addresses: A socio-technical investigation. In *Proceedings of First Conference on Email and Anti-Spam (CEAS) 2004* (Mountain View, CA, July 2004).

[22] Halverson, C. A. Activity theory and distributed cognition: Or what does cscw need to do with theories? *Comput. Supported Coop. Work 11*, 1-2 (2002), 243–267.

[23] Harms, S. K., Deogun, J., and Goddard, S. Building knowledge discovery into a geo-spatial decision support system. In *SAC '03: Proceedings of the 2003 ACM symposium on Applied computing* (New York, NY, USA, 2003), ACM Press, pp. 445–449.

[24] Harrison, J. R., and Carroll, G. R. The dynamics of cultural influence networks. *Comput. Math. Organ. Theory 8*, 1 (2002), 5–30.

[25] Hogg, T., and Adamic, L. Enhancing reputation mechanisms via online social networks. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce* (New York, NY, USA, 2004), ACM Press, pp. 236–237.

[26] Horn, D. B., Finholt, T. A., Birnholtz, J. P., Motwani, D., and Jayaraman, S. Six degrees of jonathan grudin: a social network analysis of the evolution and impact of cscw research. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work* (New York, NY, USA, 2004), ACM Press, pp. 582–591.

[27] Kapopoulos, D. G., Hatzopoulos, M., and Stamatopoulos, P. Efficient management of persistent knowledge. *J. Intell. Inf. Syst. 19*, 1 (2002), 111–134.

[28] Kempe, D., Kleinberg, J., and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2003), ACM Press, pp. 137–146.

[29] Khoussainov, R., and Kushmerick, N. Email task management: An iterative relational learning approach. In *Proceedings of Second Conference on Email and Anti-Spam CEAS 2005* (Stanford University, July 21 & 22 2005).

[30] Lavra, N., Motoda, H., and Fawcett, T. Editorial: Data mining lessons learned. *Mach. Learn. 57*, 1-2 (2004), 5–11.

[31] Liben-Nowell, D., and Kleinberg, J. The link prediction problem for social networks. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management* (New York, NY, USA, 2003), ACM Press, pp. 556–559.

[32] Lomi, A., and Pattison, P. Introduction to the cmot special issue on mathematical representations and models for the analysis of social networks within and between organizations. *Comput. Math. Organ. Theory 10*, 1 (2004), 5–15.

[33] Mařík, V., Pěchouček, M., and Štěpánková, O. Social knowledge in multi-agent systems. In *Multi-Agent Systems and Applications : 9th ECCAI Advanced Course ACAI 2001 and Agent Link's 3rd European Agent Systems Summer School, EASSS 2001* (New York, NY, USA, 2001), Springer-Verlag New York, Inc., pp. 211–245.

[34] McDonald, D. W. Recommending collaboration with social networks: a comparative evaluation. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2003), ACM Press, pp. 593–600.

[35] Mirza, B. J., Keller, B. J., and Ramakrishnan, N. Studying recommendation algorithms by graph analysis. *J. Intell. Inf. Syst. 20*, 2 (2003), 131–160.

[36] Neustaedter, C., Brush, A. B., and Smith, M. A. The social network and relationship finder: Social sorting for email triage. In *Proceedings of Second Conference on Email and Anti-Spam CEAS 2005* (Stanford University, July 21 & 22 2005).

[37] Secker, A., Freitas, A., and Timmis, J. AISEC: An Artificial Immune System for E-mail Classification. In *Proceedings of the Congress on Evolutionary Computation* (Canberra. Australia, December 2003), R. Sarker, R. Reynolds, H. Abbass, T. Kay-Chen, R. McKay, D. Essam, and T. Gedeon, Eds., IEEE, pp. 131–139.

[38] Spiegler, I. Knowledge management: a new idea or a recycled concept? *Commun. AIS 3*, 4es (2000), 2.

[39] Spiegler, I. Technology and knowledge: bridging a "generating"gap. *Inf. Manage. 40*, 6 (2003), 533–539.

[40] Stang, T., Pourbayat, F., Burgess, M., Canright, G., Engø, K., and Åsmund Weltzien. Archipelago: A network security analysis tool. In *LISA '03: Proceedings of the 17th USENIX conference on System administration* (Berkeley, CA, USA, 2003), USENIX Association, pp. 149–158.

[41] Tamma, V., and Bench-Capon, T. An ontology model to facilitate knowledge-sharing in multi-agent systems. *Knowl. Eng. Rev. 17*, 1 (2002), 41–60.

[42] Wakolbinger, T., and Nagurney, A. Dynamic supernetworks for the integration of social networks and supply chains with electronic commerce: modeling and analysis of buyer–seller relationships with computations. *Netnomics 6*, 2 (2004), 153–185.

[43] Yu, B., and Singh, M. P. Searching social networks. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems* (New York, NY, USA, 2003), ACM Press, pp. 65–72.

[44] Zack, M. H. Researching organizational systems using social network analysis. In *HICSS '00: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 7* (Washington, DC, USA, 2000), IEEE Computer Society, p. 7043.