

Knowledge Discover in Multi-relational Social Network

Ford Lumban Gaol & Belawati Widjaja

Fakultas Ilmu Komputer, Universitas Indonesia

Kampus UI Depok Jawa Barat, Indonesia

e-mail: flg50@ui.ac.id

ABSTRAK

Community mining is one of the major directions in social network analysis. Social network analysis has attracted much attention in recent years. Most of the existing methods on community mining assume that there is only one kind of relation in the network, and moreover, the mining results are independent of the users' needs or preferences. However, in reality, there exist multiple, heterogeneous social networks, each representing a particular kind of relationship, and each kind of relationship may play a distinct role in a particular task. Thus mining networks by assuming only one kind of relation may miss a lot of valuable hidden community information and may not be adaptable to the diverse information needs from different users.

In this paper, we systematically analyze the problem of mining hidden communities on heterogeneous social networks. Based on the observation that different relations have different importance with respect to a certain query, we propose a better new method for learning an optimal linear combination of these relations which can best meet the user's expectation. With the obtained relation, better performance can be achieved for community mining.

Kata kunci: Social network analysis, community mining, valuable hidden community information

I. PENDAHULUAN

Analisa Jaringan Sosial (AJS) merupakan pemadanan dan pengukuran hubungan dan arus diantara manusia, kelompok, organisasi, komputer, atau obyek pemrosesan informasi/pengetahuan. Analisa Jaringan Sosial (AJS) merupakan tema yang telah diperkenalkan oleh Milgram [6]. Berdasarkan penelitiannya, diestimasi bahwa setiap orang di dunia ini hanya memiliki enam (edges) antara yang satu dengan yang lainnya. Ada beberapa tahapan untuk investigasi dalam jaringan social dan aspek algoritmik dari jaringan social. Banyak upaya yang diterapkan untuk memperluas jaringan social

untuk berbagai tujuan [20][21], melakukan mining jaringan nilai pelanggan, dan menemukan kesukaan yang sama [11].

Penelitian terdahulu dalam bidang psikologi dan sosiologi menunjukkan kurangnya data dan focus untuk jaringan yang kecil [35]. Dengan meningkatnya data web, banyak data jaringan social yang potensial tersedia dan upaya penelitian menjadi lebih besar.

Schwartz & Wood melakukan proses penambangan hubungan social dari jaringan email logs [31]. Proyek ReferralWeb [21] mengusulkan untuk menambang jaringan social dari data web ang beragam dan menggunakannya untuk membantu pakar.

ADamic dan Adar [1] mencoba untuk menemukan interaksi social diantara orang dari informasi pada homepage mereka.

Agrawal et al. menganalisa perilaku social pada newsgroups [2].

Lebih lanjut Google's PageRank [28] & Kleigberg's HITS algorithm [22] dapat digambarkan sebagai analisis jaringan social pada penambangan komunitas web.

Makalah ini dibagi sebagai berikut, penambangan fitur, ekstraksi relasi, algoritma yang digunakan, lalu aplikasi dan kesimpulan.

II. PENAMBANGAN FITUR (FEATURE)

Sebagai bagian dari topic lanjut dalam data mining, Analisa Jaringan Sosial merupakan bagian dari penambangan semi-struktur fitur.

2.1. Penambangan Komunitas

Dengan pertumbuhan web, penambangan komunitas semakin menarik. Banyak penelitian telah dilakukan untuk proses pada komunitas web pages [16], literature ilmiah dari Web [7], dan *document citation database* [8].

Secara formal, dapat didefinisikan bahwa komunitas merupakan sekelompok obyek yang men-share beberapa sifat umum. Penambangan komunitas secara graf teoritis merupakan persoalan *graph-cut*. [9]. menggunakan konsep bipartite graph untuk

menemukan inti dari komunitas, dan kemudian diperluas untuk mendapatkan komunitas yang lain. Flake et al. menerapkan kerangka maximum-flow dan minimum cut pada penambangan komunitas [13]. Ide authority-&-hub juga digunakan dalam community mining [16][8] dan memiliki banyak perluasan [9]. Ide dari frequent itemset dalam association rule mining juga telah digunakan dalam community mining.

Baik analisa jaringan sosial dan penambangan komunitas dapat dilihat sebagai graph mining. Penambangan komunitas dapat dilihat sebagai identifikasi sub-graph. Penelitian terdahulu dalam graph mining dapat ditemukan dalam [10]. Semua teknik dalam graph mining dan penambangan komunitas berdasarkan *homogenous graph*, yakni hanya terdapat satu hubungan diantara obyek. Namun, dalam jaringan sosial sebenarnya, terdapat banyak hubungan diantara obyek. Untuk mengatasi hal ini, makalah ini memfokuskan pada penambangan komunitas multi-relasional.

2.2. Penambangan Relasional (*Relational mining*)

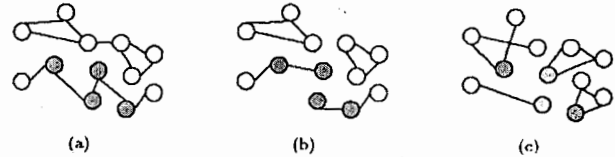
Penambangan relasional khususnya multi-relational data mining merupakan topic yang sangat menarik dan menantang pada tahun-tahun belakang ini [15]. Data mining multi-relational ditujukan untuk berhubungan dengan pencarian pengetahuan dari database relasional yang terdiri dari *multiple tables*. Hal ini mencoba untuk menganalisa data dari database multirelasional secara langsung, tanpa harus mentransfer data dalam table tunggal pertama kali.

2.3. EKSTRAKSI FITUR

Ekstraksi fitur merupakan topik terbaru yang menjadi perhatian utama dalam machine learning pada bagian klasifikasi & kluster. Ekstraksi fitur dapat dilihat sebagai menemukan kombinasi linier dari fitur asal yang dapat dengan lebih mudah menjelaskan struktur mendasar pada kumpulan data. Metode ekstraksi fitur termasuk dalamnya Prinsip Analisa Komponen [12], Analisa Diskriminan Linier [12], & *Locality Preserving Projection* [19]. Ekstraksi relasi, secara mendasar berhubungan dengan ekstraksi fitur.

III. EKSTRAKSI RELASI

Jaringan sosial mengandung relasi berganda. Relasi berbeda dimodelkan dengan graf berbeda. Graf yang berbeda ini menunjukkan hubungan diantara obyek dari berbagai obyek. Untuk persoalan penambangan komunitas, hubungan relasi ini menyediakan berbagai komunitas berbeda.



Gambar 1. Terdapat tiga relasi dalam jaringan. Empat obyek berwarna diasumsikan berada dalam komunitas yang sama, sebagaimana yang diminta pengguna..

Sebagai contoh, jaringan dalam Gambar 1 memberikan tiga relasi berbeda. Misalkan user mengasumsikan empat obyek berwarna berada dalam komunitas yang sama maka:

1. Jelasnya tiga relasi ini memiliki kepentingan yang berbeda yang direfleksikan kepada kebutuhan informasi pengguna.

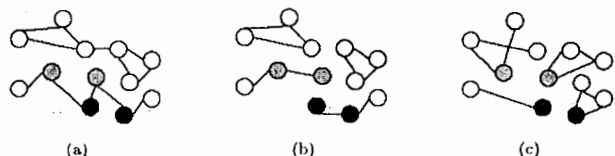
Sebagaimana yang dapat dilihat dalam relasi (a) merupakan yang terpenting dan relasi, dan (b) yang kedua serta (c) yang ketiga terpenting.

2. Dalam analisa jaringan sosial tradisional, orang tidak perlu membedakan ketiga relasi ini. Ketiga relasi ini diperlakukan sama. Sehingga cukup dikombinasikan berama-sama untuk menjelaskan struktur diantara obyek. Namun dalam relasi (c) memiliki pengaruh negative untuk hal ini. Bila dikombinasikan, dengan mudah dapat saja relasi (c) dieliminasi, sehingga relasi (a) dan (b) terpenuhi untuk digunakan untuk menemukan komunitas.

structure, yang konsisten dengan kebutuhan user.

3. Dalam analisa diatas, hubungan diantara dua obyek diasumsikan sebagai Boolean. Persoalan menjadi lebih sulit jika masing-masing edge dinyatakan dengan sebuah nilai yang menunjukkan derajat masing-masing obyek untuk saling berrelasi.

Dalam situasi yang lebih kompleks, pada Gambar 2, seorang pengguna dapat meminta query yang lebih kompleks.



Gambar 2. Diantara ketiga relasi dalam jaringan, dua obyek dengan warna lebih muda dan dua dengan warna lebih tua seharusnya berada dalam komunitas berbeda, sebagaimana yang diminta pengguna

Sebagaimana yang kita lihat dalam jaringan sosial multi-relasional, penambangan komunitas seharusnya bergantung kepada contoh user (atau informasi yang dibutuhkan). Karena permintaan pengguna (*user query*) dapat sangat fleksibel maka teknik penambangan komunitas yang lalu tidak memungkinkan untuk menjawab persoalan ini.

Persoalan ini secara matematis dimodelkan sebagai berikut. Diberikan sekumpulan obyek dan sekumpulan relasi yang dinatakan oleh sekumpulan graf $G_i(V, E_i)$, $i = 1, \dots, n$, dimana n merupakan jumlah relasi, V adalah kumpulan dari nodes (obyek) dan E_i adalah kumpulan dari edges dengan relasi i yang bersesuaian. Bobot dalam edges dapat secara alami didefinisikan sesuai dengan kekuatan relasi dari dua obyek.

Kita menggunakan M_i untuk menyatakan matriks berbobot yang bersesuaian dengan G_i , $i = 1, \dots, n$. Misalkan terdapat relasi tersembunyi yang dinyatakan oleh graf, dan $\hat{G}(V, \hat{E})$, dan \hat{M} yang dinyatakan dengan \hat{M} yang menyatakan matriks berbobot yang bersesuaian dengan \hat{G} . Diberikan sekumpulan obyek berlabel $X = [x_1, \dots, x_m]$ dan $y = [y_1, \dots, y_m]$ dimana y_j merupakan label dari x_j .

IV. ALGORITMA YANG DIGUNAKAN

Dalam bagian ini akan dibahas dua algoritma yang berdasar kepada konsep regresi [17] & min-cut [3].

4.1 Algoritma berdasar Regresi

Dasar pemikian dalam bagian ini untuk menemukan sebuah kombinasi rekas ag membolehkan hubungan dianatara contoh intra-komunitas seketat mungkin dan hubungan inter-komuntas selemah mungkin.

Relasi diantara obek berlabel adalah sebagai berikut:

$$\tilde{M}_{ij} = \begin{cases} 1, & \text{Contoh } i \text{ dan } j \text{ memiliki label} \\ & \text{yang sama} \\ 0, & \text{lainnya} \end{cases}$$

dimana \tilde{M} merupakan matriks.

Ketika matriks relasi terbentuk, maka ditujukan untuk menemukan kombinasi linier dari relasi yang ada untuk secara optimal menghampiri target relasi dalam norm L_2 . Sering kali pengguna tidak yakin jika dua obyek milik dua komunitas yang sama dan hana menyediakan kemungkinan bahwa dua obyek berada dalam komunitas yang sama.

$\tilde{M}_{ij} = \text{Prob}(x_i \& x_j \text{ berada pada komunitas yang sama})$

$$\mathbf{a} = [a_1, a_2, \dots, a_n]^T \in R^n$$

menyatakan kombinasi koefisien untuk relasi yang berbeda.

Persoalan hampiran dapat dikarakteristikan dengan menyelesaikan persoalan optimisasi:

$$\mathbf{a}^{opt} = \arg \min_{\mathbf{a}} \left\| \tilde{M} - \sum_{i=1}^n a_i M_i \right\|^2$$

Ini dapat dituliskan sebagai bentuk vector. Karena matriks $M_{m \times m}$ adalah simetrik. Dapat digunakan vector dimensi $m(m-1)/2$ untuk menyatakannya. Dapat disederhanakan menjadi

$$\mathbf{a}^{opt} = \arg \min_{\mathbf{a}} \left\| \tilde{\mathbf{v}} - \sum_{i=1}^n a_i \mathbf{v}_i \right\|^2$$

Persamaan ini, merupakan persoalan regresi [17]. Dari sudut pandang ini, hubungan persamaan ini diinterpretasikan sebagai persoalan prediksi. Bila koefisien kombinasi dapat ditemukan maka hubungan kekuatan diantara pasangan obyek dapat diprediksi. Dalam aplikasi nyata, pengguna tidak harus menyatakan hubungan diantara pasangan obyek. Sehingga, vektor \mathbf{v} tidak harus dalam dimenis $m(m-1)/2$. Diasumsikan bahwa \mathbf{v} dalam k -dimensi.

Asumsikan untuk kasus berikut

$$\sum_{i=1}^n a_i \mathbf{v}_i = \tilde{\mathbf{v}}$$

Definisikan:

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$$

Sehingga dapat dituliskan menjadi

$$V\mathbf{a} = \tilde{\mathbf{v}}$$

Misalkan *rank* dari *V* adalah $\min(k, n)$. Berikut adalah hal-hal yang perlu diperhatikan:

- bila $k < n$, solusi persamaan diatas dalam bentuk ruang vektor dimensi $(n - k)$
- bila $k = n$, terdapat solusi unik untuk persamaan diatas.
- bila $k > n$, tidak terdapat solusi untuk persamaan tersebut.

Dalam dua kasus pertama, diperoleh solusi dengan pemadanan yang sempurna (minimisasi sama dengan nol). Perhatikan bahwa *k* merupakan kuantitas informasi yang ingin diketahui oleh user, *k* adalah kecil bila query yang diminta pengguna adalah sederhana. Untuk query yang kompleks, maka *k* besar dari *n*. Maka solusi hana bisa dilakukan dengan menurunkan fungsi obektifnya ke nol.

$$\frac{\partial \|\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i\|^2}{\partial a_i} = 0 \quad \text{for } i = 1, \dots, n$$

Penyederhanaan secara aljabar diperoleh:

$$\begin{aligned} & \frac{\partial \|\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i\|^2}{\partial a_i} = 0 \\ & = \frac{\partial [(\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)^T (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)]}{\partial a_i} = 0 \\ & = \frac{\partial (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)^T}{\partial a_i} (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i) + (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)^T \frac{\partial (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)}{\partial a_i} = 0 \\ & = 2 \left\{ \frac{\partial (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i)^T}{\partial a_i} \right\}^T (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i) = 0 \\ & = \mathbf{v}_i^T (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i) = 0 \quad \text{for } i = 1, \dots, n \\ & = \mathbf{1}^T (\tilde{\mathbf{v}} - \sum_i^n a_i \mathbf{v}_i) = 0 \\ & = \mathbf{1}^T (\tilde{\mathbf{v}} - V\mathbf{a}) = 0 \\ & = \mathbf{1}^T V\mathbf{a} = \mathbf{1}^T \tilde{\mathbf{v}} \end{aligned}$$

4.2. Algoritma Min-Cut

Di bagian 4.1, telah dijelaskan metode yang umum, untuk ekstraksi relasi yang tersembunyi berdasar model regresi. Namun, metode ini gagal bila contoh dari pengguna berada dalam komunitas yang kecil, atau hanya satu komunitas. Contoh, misalkan *query user* (o1, o2, o3, o4, o5), yang berada dalam komunitas sama,

	<i>o</i> ₁	<i>o</i> ₂	<i>o</i> ₃	<i>o</i> ₄	<i>o</i> ₅
<i>o</i> ₁	*	0.9	0.8	1	0.1
<i>o</i> ₂	0.9	*	1	0.9	0.1
<i>o</i> ₃	0.8	1	*	1	0.1
<i>o</i> ₄	1	0.9	1	*	0.1
<i>o</i> ₅	0.1	0.1	0.1	0.1	*

(a) Relation *M*₁

	<i>o</i> ₁	<i>o</i> ₂	<i>o</i> ₃	<i>o</i> ₄	<i>o</i> ₅
<i>o</i> ₁	*	0.4	0.4	0.5	0.5
<i>o</i> ₂	0.4	*	0.3	0.2	0.4
<i>o</i> ₃	0.4	0.3	*	0.3	0.3
<i>o</i> ₄	0.5	0.2	0.3	*	0.4
<i>o</i> ₅	0.5	0.4	0.3	0.4	*

(b) Relation *M*₂

Dalam dua relasi berikut, model regresi akan memilih *M*₁, karena konektivitas tertinggi diantara *o*₁, *o*₂, *o*₃, *o*₄ mencapai square error untuk target relasi. Namun, dalam relasi *M*₁, konektivitas diantara *o*₅ dan empat contoh lainnya adalah sangat lemah. Sebagaimana yang dilihat bahwa *M*₂ jauh lebih seragam dibandingkan *M*₁. Maka *M*₂ merupakan pilihan yang lebih baik untuk query user. Namun secara regresi *M*₂ jauh lebih square error-na jauh lebih besar daripada *M*₁, sehingga hal ini contoh kasus yang gagal untuk model regresi.

Untuk mengatasi masalah komunitas tunggal, akan dilihat hubungan terlemah dalam mengekstrak relasi. Menggunakan teori graf maka nilai minimum cut dalam teori graf dapat digunakan untuk mengevaluasi kerapatan dari graf. Misalkan *G* merupakan graf dengan matriks bobot *M*. Asumsikan *m* jumlah verteks.

Sebuah cut dalam graf didefinisikan sebagai kumpulan edges yang memisahkan verteks kedalam dua kelompok terpisah A dan B sehingga $A \cap B =$

ϕ and $A \cup B = G$. Maka nilai dari cut adalah sebagai berikut:

$$cut(G) = \sum_{i \in A} \sum_{j \in B} M(i, j)$$

Adalah mudah untuk melihat bahwa terdapat $2^m - 2$ cut yang berbeda. Misalkan

$cut_k(G) = (A_k, B_k)$ sebagai k-th cut.

Minimum cut didefinisikan sebagai:

$$mincut(G) = \min_k \{cut_k(G)\}$$

Jika graf dapat dengan mudah di-cut menjadi dua subgraf, bila memiliki nilai cut yang kecil. Untuk kasus ekstrem nilai cut untuk *disconnected* graf adalah nol. Secara alami, relasi ekstraksi optimal harus memiliki nilai minimumcut yang besar.

Sehingga untuk komunitas tunggal, akan dicoba untuk memaksimalkan nilai minimum cut.

Misalkan $G_i, i = 1, \dots, n$, menyatakan relasi graf yang ada yang didefinisikan oleh user. Misalkan $a = [a_1, a_2, \dots, a_n]^T \in R^n$ sebagai kombinasi koefisien untuk graf yang berbeda.

Fungsi obyektif dituliskan sebagai berikut;

$$a^{opt} = \arg \max_a \{mincut(\sum_{i=1}^n a_i G_i)\}$$

Secara umum, persoalan *minimum cut* adalah NP-hard problem. Maka persoalan optimisasi diatas tidak dapat diselesaikan dengan mudah. Namun berdasar kenyataan, jumlah contoh yang disediakan oleh user biasanya kecil, dimana m kurang dari 10. Maka bisa diselesaikan dengan pemrograman linier sbb:

$$\begin{aligned} mincut(G) &= \min_{1 \leq k \leq 2^m - 2} \{cut_k(G)\} \\ &= \min_{1 \leq k \leq 2^m - 2} \left\{ \sum_{i \in A(k)} \sum_{j \in B(k)} M(i, j) \right\} \\ &= \min_{1 \leq k \leq 2^m - 2} \left\{ \sum_{i \in A(k)} \sum_{j \in B(k)} \left(\sum_{l=1}^n a_l M_l(i, j) \right) \right\} \\ &= \min_{1 \leq k \leq 2^m - 2} \left\{ \sum_{l=1}^n a_l \left(\sum_{i \in A(k)} \sum_{j \in B(k)} M_l(i, j) \right) \right\} \\ &= \min_{1 \leq k \leq 2^m - 2} \left\{ \sum_{l=1}^n a_l \cdot cut_k(G_l) \right\} \end{aligned}$$

Misalkan $v = mincut(G)$. Persoalan optimisasi diatas dapat direduksi untuk persoalan pemrograman linier:

$$\begin{aligned} &\max v \\ \text{st. } &\sum_{l=1}^n a_l \cdot cut_k(G_l) - v \geq 0, \quad (1 \leq k \leq 2^m - 2) \quad (*) \\ &\sum_{l=1}^n a_l = 1 \\ &a_l \geq 0, \quad (1 \leq l \leq n) \end{aligned}$$

V. APLIKASI POTENSIAL

Dengan berkembangnya internet, jaringan sosial yang heterogen menjadi hal yang umum. Pada bagian ini akan dibahas beberapa aplikasi yang potensial dari ekstraksi relasi. Berdasar terhadap pengamatan selama ini, jaringan sosial heterogen dalam dapat dikategorikan sebagai berikut.

Social Community Websites. *Social community websites* tidak begitu dikenal pada masa yang lalu. Contoh dari komunitas ini adalah www.orkut.com & www.xanga.com. Berdasar pengamatan sampai dengan Dec 2006, www.orkut.com telah memiliki populasi lebih dari 3.2 juta pengguna yang terregistrasi. Pada website seperti itu, seorang pengguna akan terregister oleh referensi temannya. Maka seluruh pengguna yang terregister membentuk sebuah graf sosial yang besar. User dapat membentuk link terhadap beragam teman. Teman-teman mereka biasanya men-share berbagai komunitas tertentu. Sebagai contoh, seseorang memiliki teman pada waktu SLTA, beberapa teman pada waktu dia mahasiswa, teman kerjanya dan

beberapa diantaranya memiliki hobi yang sama seperti panjat tebing atau berenang, atau lainnya.

Forums & Newsgroups. *Forums & newsgroups* adalah tempat yang populer dimana banyak orang membahas beberapa topic yang khusus dengan yang lainnya. Mereka juga merepresentasikan berbagai topic yang khusus. Forum yang cukup terkenal adalah www.craigslist.org, dimana banyak orang berpartisipasi dalam diskusi dan hal-hal lainnya.

Instant Messengers & Emails. *Instant Messengers (IMs) & Emails* adalah dua media yang digunakan banyak orang untuk berkomunikasi dengan yang lainnya. Instant Messenger merupakan media yang paling banyak digunakan saat ini sebagai alat bantu untuk memberikan pesan singkat kepada orang diseluruh dunia. MSN Messenger. Banyak orang yang terdaftar sebagai IM dan selalu dikelola dalam daftar IMS. Secara default kelompok men-set up MSN Messenger termasuk dalamnya teman, saudara, dll. Kelompok ini menyediakan informasi bermanfaat dalam hubungan semantik. Informasi demikian dapat diutilisasi dalam analisa jaringan multi-relasional. E-mail merupakan contoh data lainnya yang digunakan untuk menambang hubungan diantara orang. Beberapa penelitian terdahulu telah dimulai [13]. Jaringan sosial multi-relasional menyediakan kesempatan besar untuk algoritma ekstraksi relasi untuk eksplorasi aplikasi sebagaimana yang dinyatakan sebagai berikut:

1. *Saran teman.* Banyak orang yang bergabung dalam komunitas sosial berbasis web, yang bertujuan untuk menemukan pasangan hidup, teman maya, atau rekan bisnis. Diberikan derajat besar untuk masing-masing node dalam jaringan sosial dan secara eksponensial meningkat kandidat seiring dengan bertambahnya panjangnya link teman yang secara otomatis membutuhkan waktu yang lebih besar untuk menemukan orang yang dicari. Model yang digunakan dalam makalah ini menyediakan solusi yang realistis, Sistem mengumpulkan contoh-contoh dari input user dan melakukan training terhadap bobot masing-masing jaringan yang diinginkan. Kombinasi data yang di-train, dapat digunakan untuk melakukan saran teman.

2. *Targeted marketing.* *Targeted marketing* seperti *viral marketing & content-based advertisement*, adalah populer baik dalam bidang penelitian dan bisnis. Dengan menggunakan penambangan jaringan sosial multi-relasional, maka dapat memberikan saran yang lebih akurat dalam *targeted marketing*. Sebagai contoh, seseorang yang membeli berbagai

produk atau berpartisipasi dalam berbagai aktivitas dapat membentuk jaringan sosial. Ketika sebuah produk diluncurkan ke pasar maka dengan mudah untuk menemukan pelanggan potensial berdasar kepada evaluasi yang mendalam terhadap jaringan sosial multiple.,

3. *Network prediction.* Pemodelan dari ealosu jaringan sosial secara aktif telah dibahas secara mendasar tahun belakangan ini. Namun, seluruh pendekatan yang digunakan masih berupa jaringan sosial tunggal, Kenyataannya, jaringan sosial ternyata memiliki pengaruh yang tidak independen. Sebagai contoh, hubungan perkawinan dapat secara kuat berhubungan dengan jaringan sosial seperti bekas teman sekolah, atau kesamaan hobi. Namun, bobot dari jaringan ini dapat berubah bila waktu atau lokasi berubah. Dapat juga diobservasi digunakan untuk memprediksi pernikahan yang ternyata dipengaruhi oleh jaringan sosial lainnya. Ide utama dari aplikasi ini berasal dari persoalan yang sama: bagaimana menemukan jaringan relasi tersembunyi tingkat tinggi bila diberikan jaringan sosial tingkat rendah.

VI. PEMBAHASAN

Karena penambangan komunitas tersembunyi dalam jaringan heterogen merupakan sebuah daerah penelitian yang sangat luas, maka banyak pertanyaan terbuka yang dapat dibahas disini. Pada bagian ini akan ditekankan kepada *problem solving*. Bagian awal akan ditekankan kepada kompleksitas dan kombinasi dari analisa jaringan sosial. Kita setuju bahwa jaringan sosial berganda merupakan bentuk dari interrelasi graf, dan dengan semakin besarnya jumlah data maka tantangan yang dihadapi seseorang untuk memperoleh gambaran lengkap pada jaringan sosial yang lengkap sangat bergantung kepada keseimbangan dari jaringan sosial ganda tersebut dan juga bergantung kepada query dari pengguna. Namun, jaringan berganda yang ada adalah tidak tepat langsung digabung untuk menghilangkan kompleksitasnya sehingga memudahkan untuk menganalisisnya. Maka, dipercaya bahwa pengembangan algoritma penambangan jaringan sosial berganda untuk secara dinamis mengkombinasikan jaringan yang relevan untuk membentuk "jaringan maya berdasar kepada pengguna" merupakan contoh dalam metodologi *problem solving*.

Kedua, karena sulit untuk seorang pengguna untuk melihat secara komprehensif keseluruhan jaringan sosial secara lengkap.

Berdasar pengalaman walaupun sulit untuk mengumpulkan keseluruhan jaringan berganda, seorang user biasanya memiliki kumpulan contoh yang lengkap (yang mempengaruhi peneliti, bintang film, perusahaan besar, dll). Berdasar pengalaman terhadap DBLP ternyata banyak proses learning yang dapat dilakukan untuk menghasilkan pengetahuan terhadap publikasi yang dilakukan.

Ketiga, bagaimana memperoleh jawaban yang diperoleh dari analisa jaringan karena banyaknya jaringan tersembunyi yang merupakan matriks berbobot sebagai kombinasi dari jaringan berganda yang sulit untuk dipahami perbedaan hasil yang minor. Esensi terpenting pada dasarnya pada fakta terbaru yang diturunkan dari jaringan tersembunyi dan hubungannya terhadap tingkat yang bersesuaian. Hal ini ditunjukkan pencarian web oleh Google berdasar kata kunci. Hal ini tidak begitu krusial karena pemahaman dari ranking bobot web ternyata tidak begitu signifikan karena munculnya spam yang secara otomatis meningkatkan ranking suatu website.

VII. KESIMPULAN

Berbeda dengan analisa jaringan sosial yang biasa dilakukan banyak orang, maka dalam makalah ini terdapat jaringan sosial heterogen berganda dan kombinasi yang rumit dari jaringan sosial heterogen tersebut dapat membangkitkan hubungan yang baru yang penting yang memenuhi kebutuhan pengguna berdasar informasi dari user. Maka pendekatan yang dilakukan dalam makalah ini merupakan peningkatan terbaru terhadap metodologi tradisional, jaringan tunggal dimana pengguna tidak diikutsertakan, menjadi jaringan multijaringan dan berdasar kepada analisa query pengguna. Argumen atau latar belakang pemilihan metodologi ini muncul karena kenyataan dalam dunia nyata memang seperti hal, dimana aktivitas seseorang dalam bidang tertentu akan berpengaruh dalam bidang lainnya.

Berdasar metodologi tersebut dikembangkan metodologi baru untuk ekstraksi relasi dan mengusulkan dua algoritma dalam situasi berbeda. Dengan ekstraksi relasi berdasar ketergantungan query. Dalam diskusi juga ditunjukkan bahwa ekstraksi relasi berdasar query dan penambahan

komunitas akan memberikan aplikasi potensial dalam analisa jaringan sosial. Studi yang dilakukan bahwa metodologi ini sangat menjanjikan dan dapat diterapkan dalam bidang lainnya.

Ada beberapa bidang yang mungkin dapat ditelusuri lebih lanjut.

Pertama, pendekatan menggunakan analisa regresi berdasar graf. Terdapat banyak potensi yang mungkin bisa diterapkan dan dieksplorasi serta dibandingkan dengan metode ini.

Kedua, algoritma ekstraksi relasi telah membuat banyak penyederhanaan dalam analisa. Secara umum, *link* dalam jaringan sosial dan antar jaringan dapat membawa bobot berbeda.

Ketiga, dalam model query hanya satu kelompok nodes (seperti peneliti). Query yang lebih umum dapat lebih ditingkatkan. Sebagai contoh, seseorang yang ingin mengetahui mereka yang hadir dalam konferensi namun tidak pernah muncul sebagai penulis kedua dalam konferensi. Ini dapat membantu untuk menemukan *reviewer* untuk penyampaian kalah dalam konferensi tersebut.

Is-isu ini merupakan kesempatan dalam pengembangan metode ini dikemudian hari.

DAFTAR PUSTAKA

- [1] L. A. Adamic and E. Adar. Friends and neighbors on the web. Technical report, Xerox Parc, 2002.
- [2] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu. Mining newsgroups using networks arising from social behavior. In Proceedings of 12th International World Wide Web Conference, 2003.
- [3] M. Bazaraa, J. Jarvis, and H. Sherali. Linear Programming and Network Flows. Wiley, 3rd edition, 2004.
- [4] A. Bjorck. Numerical Methods for Least Squares Problems. SIAM, 1996.
- [5] D. Cai, X. He, J.-R. Wen, and W.-Y. Ma. Block-level link analysis. In Proceedings of the ACM SIGIR'2004, 2004.
- [6] S. Chakrabarti. Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction. In Proceedings of the 10th International World Wide Web Conference, 2001.
- [7] S. Chakrabarti, M. van den Berg, and B. Dom. Focused crawling: a new approach to topic-specific web resource discovery. In Proceedings of The 8th International World Wide Web Conference, 1999.
- [8] C. Chen and L. Carr. Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In Proceedings of the 10th ACM Conference on Hypertext and hypermedia, 1999.

- [9] D. Cohn and H. Chang. Learning to probabilistically identify authoritative documents. In Proceedings of the 17th International Conference on Machine Learning, 2000.
- [10] D. J. Cook and L. B. Holder. Graph-based data mining. *IEEE Intelligent Systems*, 15(2):32–41, 2000.
- [11] P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 57–66. ACM Press, 2001.
- [12] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, Hoboken, NJ, 2nd edition, 2000.
- [13] G. W. Flake, S. Lawrence, and C. L. Giles. Efficient identification of web communities. In Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000), 2000.
- [14] G. W. Flake, S. Lawrence, C. L. Giles, and F. Coetzee. Self-organization of the web and identification of communities. *IEEE Computer*, 35(3):66–71, 2002.
- [15] J. Gehrke, R. Ramakrishnan, and V. G. Rainforest. A framework for fast decision tree construction of large datasets. In Proc. 1998 Int. Conf. Very Large Data Bases (VLDB'98), 1998.
- [16] D. Gibson, J. Kleinberg, and P. Raghavan. Inferring web communities from link topology. In Proceedings of the 9th ACM Conference on Hypertext and Hypermedia, 1998.
- [17] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [18] T. Haveliwala. Topic-sensitive pagerank. In Proceedings of the 11th International World Wide Web Conference, May 2002.
- [19] X. He and P. Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems*, 16, 2003.