

動的ネットワークのマイニングと企業価値の予測

Mining Longitudinal Networks to Predict Company Value

金英子*1
Yingzi JinChing-Yung Lin*2
Ching-Yung Lin松尾豊*3
Yutaka Matsuo石塚満*4
Mitsuru Ishizuka*1, *3 東京大学大学院 工学系研究科
School of Engineering, The University of Tokyo*2 IBM トーマス・J・ワトソン研究所
IBM T.J. Watson Research Center*4 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

Real-world social networks are dynamic in nature. Companies continue to collaborate, align strategically, acquire, and merge over time, and receive positive/negative impact from other companies. Consequently, their performance changes with time. A person who can understand what types of network changes affect a company's value can predict the future value of the company, grasp industry innovations, and make business more successful. However, it often requires continuous records of relational changes, which are often difficult to track for companies, and the models of mining longitudinal network are quite complicated. In this study, we developed algorithms and a system to infer large-scale evolutionary company networks from public news during 1981–2009. Then, based on how networks change over time, as well as the financial information of the companies, we predicted company profit growth.

1. はじめに

社会ネットワーク分析では、アクターのもつ関係性とその構造に着目することで、アクターの行為やパフォーマンスを解釈する [Wasserman 94]. 例えば、ある企業は積極的にたくさんの企業と連携して事業を進めているが、他の企業は一定の企業としか提携しない；ある企業の提携先企業は互いに連携しているが、他の企業の提携先企業はほとんど関係しない。これらの関係やその構造は、企業の行為や価値に影響を与える。従来の研究では、次数中心性や媒介中心性などの指標を用いてアクターのパフォーマンスを計量化している [Uzzi 97].

一方で、実世界のネットワークの本質は動的に変化する。企業は、他の企業と提携・合併・買収などを続けることで、これらの関係から正と負の影響を受けている。その結果、企業の価値は変わっていく。この企業のネットワークの変化と企業価値の変化のメカニズムを分析することで、企業の戦略や業界の動向を把握することができる。社会学では、動的ネットワークを分析することで、ネットワークの進化、信念の形成、友達の形成などを解析している [Doreian and Stokman 97, McCulloh and Carley 09].

本研究では、Web上に公開されているニュース記事から、変化する企業間関係のネットワークを自動的に抽出する。そして、この動的ネットワークから企業の価値に影響を与える構造的要素を抽出し、企業の価値を予測する。本研究の目的は、ネットワークの変化と価値変化の間のメカニズムを解明することである。これには、企業同士の関係性をどのように抽出するか、企業の関係構造の変化から企業価値の変化をどのように予測するか、という2つの課題が存在する。最初の課題について、我々は公開されているニュース記事から、企業の動的ネットワークを抽出するアルゴリズムを提案する。特に、企業が他社から受けるインパクトに注目する。インパクトの強さは、対象企業が他の企業とどれだけ多くの重要な記事・文脈の

中で共起するかの情報を利用する。そして、次の課題については、抽出した変化する企業間関係のネットワークから、企業が受けるネットワーク効果を定義して抽出し、それらの特微量として学習することで、企業の価値を予測する。

2. ニュース記事から動的ネットワークの抽出

本研究では、ニューヨーク・タイムズ (1981年～2009年) を用いて、企業の年ごとの動的ネットワークを抽出する。7,594企業がインデックスされている。特に対象企業に関しては、価値の変化とネットワーク変化を結び付けて分析するために、アメリカフォーチュン誌が毎年発表するフォーチュン500^{*1}に3回以上ランキングされている企業として、それらの企業の毎年の利益値 (Profit) を企業の価値とする。

Web上に公開されているニュース記事から企業の共起関係のネットワークを抽出する研究がある [Bernstein et al. 02]. そこで示しているように、長期的視点で企業の価値を評価するには、その企業に関する広い期間のニュースを集めて、企業の活躍ぶりを分析する必要がある。例えば、ニューヨーク・タイムズの2009年には、IBM社に関する記事がおおよそ300件ある^{*2}。これらの記事をすべて読むことができれば、IBM社がこの1年の間、どの企業からどれだけのインパクトを受けて、その価値がどう変わったかが分かる。本研究では、インパクト関係を次のように定義する。ある企業が頻繁に対象企業と重要な記事で言及され、頻繁に対象企業と重要な文節で記述されていれば、この企業は対象企業に強いインパクトを与える。

本研究では、文書と文の共起情報を用いてその頻度を計り、各文書と文に重要度を与えることでインパクト関係を計算する。具体的に、各対象企業 x に対して、各候補企業 $y \in Y$ が一定期間に与えるインパクトの強さを計算する。ここで候補企業は、一定期間内で企業 x に関するニュース記事に現れた企業とする。まず、各候補企業 y に対して、 x と一定期間内に共

連絡先: 金英子, 東京大学大学院工学系研究科, 〒113-8656 東京都本郷7-3-1, TEL/FAX: 03-5841-7672, Email: eikokin@biz-model.t.u-tokyo.ac.jp

*1 <http://money.cnn.com/magazines/fortune/fortune500/2010/>
*2 (I.B.M.として277件, International Business Machinesとして84件)

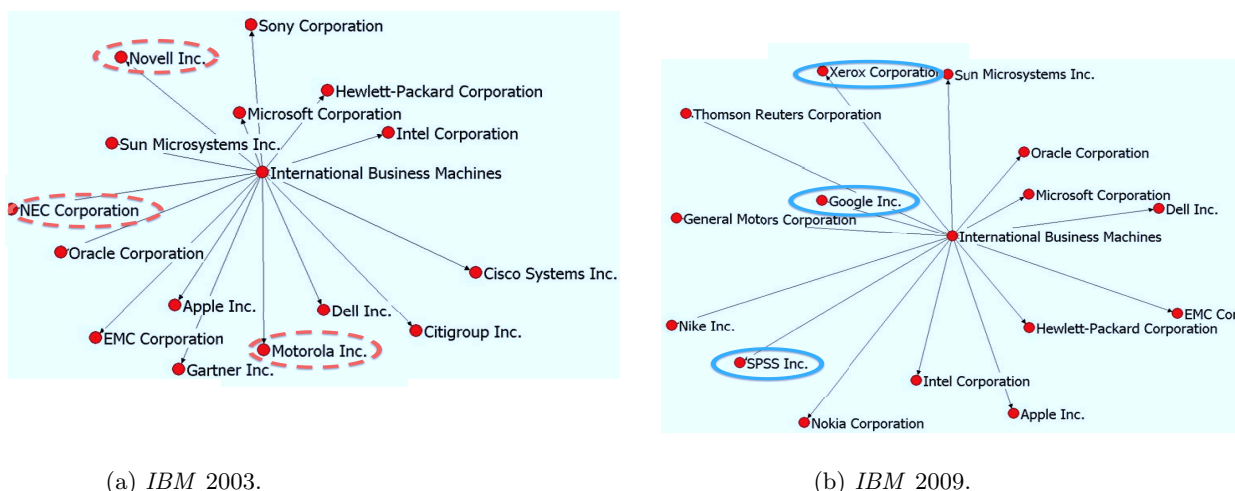


図 1: Evolution of Networks in Different Years.

表 1: Example of generic relation extraction for IBM in 2009.

r	name	score	Examples of documents and sentences.
1	Microsoft	85.85	<i>I.B.M. - Microsoft (55 articles, 264 sentences, score=85.85455)</i> http://www.nytimes.com/2009/03/06/business/06layoffs.html
2	Oracle	65.49	- Two days after <u>I.B.M.</u> 's report, <u>Microsoft</u> said that its quarterly profits were disappointing.
3	Google	57.75	http://www.nytimes.com/2009/01/31/business/31nocera.html
4	HP	50.70	- <u>Caterpillar</u> , <u>Kodak</u> , <u>Home Depot</u> , <u>I.B.M.</u> , even mighty <u>Microsoft</u> : they are all cutting jobs.
5	Intel	48.52	<i>I.B.M. - SPSS (1 articles, 9 sentences, score =13.675)</i>
6	Dell	32.75	http://www.nytimes.com/2009/07/29/technology/companies/29ibm.html
7	Sun	29.45	- <u>I.B.M.</u> to Buy <u>SPSS</u> , a Maker of Business Software
8	EMC	15.16	- <u>I.B.M.</u> 's \$50-a-share cash offer is a premium of more than 40 percent over <u>SPSS</u> 's closing stock price...
9	Apple	14.65	<i>I.B.M. - Nike (4 articles, 9 sentences, score =8.212)</i>
10	SPSS	13.67	http://www.nytimes.com/2009/01/22/business/22pepsi.html
11	GM	13.18	- The list of companies that have taken steps to reduce carbon emissions includes <u>I.B.M.</u> ,
12	Xerox	12.13	<u>Nike</u> , <u>Coca-Cola</u> and <u>BP</u> , the oil giant.
13	Nokia	8.95	http://www.nytimes.com/2009/11/01/business/01proto.html
14	Nike	8.21	- Others are water-based shoe adhesives from <u>Nike</u> and a packing insert from <u>I.B.M.</u>
15	Thomson Reuters	7.78	

起した文書集合 $D_{x,y}^t$ と文集合 $S_{x,y}^t$ を集める。次に、これらの文書の重み $w_d(i)$ と頻度、および文の重み $w_s(j)$ と頻度の線形和を取ることで、 y が x に与えるインパクトを計算する。

$$score_x(y) = a \cdot \sum_{i \in D_{x,y}^t} w_d(i) + b \cdot \sum_{j \in S_{x,y}^t} w_s(j). \quad (1)$$

ここで、各共起する文書と文の重みは下記のように定義する。

$$w_d(i) = \log\left(1 + \frac{1}{|Y'|} + \frac{tf_x}{\sum_{y \in \{x, Y'\}} tf_y}\right) \quad (2)$$

$$w_s(j) = \log\left(1 + \frac{1}{|Y''|}\right) \quad (3)$$

ここで、 Y' と Y'' はそれぞれ文書 i と文 j に出現する企業名の集合、 $|Y'|$ と $|Y''|$ はその数を表す。 tf_y は文書 i における y の出現数を指す。直感的には、企業の名前をたくさん含む文書は、そうでない文書より重要度が下がる。そして、 x と y だけを言及している文の重要度は高く、たくさんの企業名をリストしている文の重要度は低い。定数 a と b は、文書と文の重要度の間のトレードオフを表す。経験的に $a=1, b=5$ にした*3。

*3 学習データに対して、複数セットの a と b を用いて生成したネッ

最終的に、企業同士の年ごとネットワークを抽出することができる。図 1 は、IBM 社の 2003 年と 2009 年のエゴ・ネットワーク (の一部) を表す。ここから、IBM 社のネットワークの変化を見ることができる。例えば、Motorola, Novell, NEC 社は、2003 年には IBM 社に対して高いインパクトを与えていたが、2009 年には上位からなくなり、その代わりに Google, SPSS, Xerox 社が強いインパクトを与えていることが分かる。それ以外の上位の企業も、少なからずインパクトが変わっていることが分かる。表 1 に、2009 年に IBM 社に影響を与えた上位 15 社の企業と、その記事で出現する共起文 (の一部) を列挙する。(ページ数制限のため、ここではその解釈を省く。)

抽出されたネットワークは重み付き有向ネットワークである。つまり、企業 a が企業 b に与えるインパクトと、企業 b が企業 a に与えるインパクトは異なる。なお、本研究ではインパクト関係を、ポジティブとネガティブ (すなわち、正と負の関係) に分けて考慮しない。あくまでも、ネットワークの構造とその変化から正と負の影響を受けることに着目する。

トワークをユーザに見せて決めている。

3. ネットワークに基づいた企業価値の予測

企業同士の動的ネットワークを構築した後、本節ではその変化のメカニズムを分析することで、企業の価値を予測する。

3.1 ネットワーク効果

まず、各対象ノード x が受けるネットワーク効果を調べる。現時点ネットワーク効果、過去のネットワーク効果、ネットワーク変化による効果を合わせて、ベクトル \mathbf{F}_x^t を用いて表す。

x が受ける現時点ネットワーク効果 (\mathbf{F}_x^t とする) は、唐門ら [Karamon et al.] の研究に基づいて計算する。まず、 x に影響を及ぼすノードセット N_x を決める。次に、 N_x に対して3つのタイプのノードペアを定義する： $\langle x, i \rangle$ (ここで、 $i \in N_x$)、 $\langle i, j \rangle$ (ここで、 $i \in N_x, j \in N_x, i \neq j$)、 $\langle i, k \rangle$ (ここで、 $i \in N_x, k \in V$)。そして、各ノードペアに対して3つの種類の演算式を定義する：*connectivity* $\beta(i, j)$ (i.e., 到達可能なら1を、そうでなければ0を返す)、*distance* $\mu(i, j)$ (i.e., i と j の最短距離を返す)、*betweenness* $\zeta^x(i, j)$ (i.e., i と j の最短パスに x を含むなら1を、そうでなければ0を返す)。最終的に、 N_x の x に関する6つのネットワーク効果を計算する。

- $\sum_{i \in N_x} \beta(x, i) / (|V| - 1)$: x のすべての隣接ノードの数
- $\sum_{i \in N_x} \mu(x, i) / (|V| - 1)$: x とすべてのノードとの距離の和
- $\sum_{k \in V} \beta(i, k) / (|V| - 1)$: x と関係するすべてのノードの数
- $\sum_{i, j \in N_x} \beta(i, j) / (|V| - 1)(|V| - 2)$: x に関係するノード同士の繋がりの数
- $\sum_{i, j \in N_x} \mu(i, j) / (|V| - 1)(|V| - 2)$: x に関係するノード同士の距離の和
- $\sum_{i, j \in N_x} \zeta(i, j) / (|V| - 1)(|V| - 2)$: 最短パスが x を経由するノードペアの数

ここで、ノードセット N_x について、 x に直接繋がっている隣接ノードセット L_x 、間接的に繋がっている到達可能なノードセット G_x に分けて、その影響を調べる。さらに、 L_x と G_x セットからの影響の割合を考慮する。例えば、 x と L_x 、 x と G_x の連結数の割合は、 x に関係する企業がどれだけ x の周りに集中しているかの具合を表す。次に、重み付き有向ネットワークに対して、一部の情報を減らすことにより、異なる種類のネットワークを得ることができる。例えば、方向だけを考慮し重みを無視した場合、Facebook や linkedIn のような、友人関係のネットワークが得られる。つまり、誰と誰が友達であるかだけを考慮し、どれだけ強い関係であるかは無視する。従って、対象ノード x の現時点ネットワーク効果は、直接関係と間接関係、その割合、さらに4つの種類のネットワーク (重み付き有向、バイナリ有向、重み付き無向、バイナリ無向) を考慮することで、 $72 = (3 \times 4 \times 6)$ 次元のベクトルになり、ここでは $\mathbf{F}_x^t = F(N_x, d, v, t)$ と表す。ここで、 $N_x \in \{L_x, G_x, L_x/G_x\}$ 、 $(v, d) \in \{(0, 1)\} \times \{0, 1\}$ 、そして $t \in T$ 。

年ごとの記事を対象に現時点ネットワーク効果を計算したあとは、時間をずらすことで過去のネットワーク効果も計算することができる。すなわち、 $\mathbf{F}_x^H = \{\mathbf{F}_x^{t-1}, \mathbf{F}_x^{t-2}, \dots, \mathbf{F}_x^{t-w}\}$ 。ここで \mathbf{F}_x^{t-w} は、 w 年前のネットワーク効果を表す。さらに、現在と過去のネットワーク変化による効果も計算することができる。すなわち、 $\Delta \mathbf{F}_x^H = \{\Delta \mathbf{F}_x^{t-1}, \Delta \mathbf{F}_x^{t-2}, \dots, \Delta \mathbf{F}_x^{t-w}\}$ 。例えば、 x の去年と今年の隣接ノード数の変化など。

最終的に、年ごとに各対象企業に対して、72次元の現時点ネットワークの効果 \mathbf{F}_x^t 、 $72 \times \text{window size}$ 次元の過去ネット

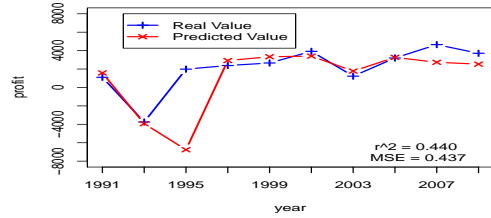


図 2: Prediction of the mean profits of 20 Fortune companies.

ワークの効果 \mathbf{F}_x^H 、および $72 \times \text{delta size}$ 次元のネットワーク変化の効果 $\Delta \mathbf{F}_x^H$ のベクトルが得られる。さらに、予測モデルでは、去年の売上げ、3年前の売上げなど、企業の属性データも組み合わせることができる。

3.2 予測モデル

対象ノード x の動的ネットワークの各効果を計量化したあと、それらを特徴量として学習することで、企業の未来の価値を予測することができる。

$$y_x^{t'} = f(\mathbf{F}_x^t, \beta) = \sum_k f_k \beta_k \quad (4)$$

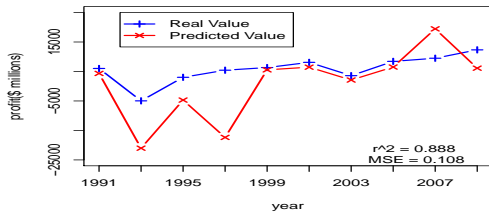
ここで、 $t' > t$ であり、 f_k は k 番目のネットワーク効果を表し、 β_k はその重要度を示す。予測モデルは、観測データから未知のパラメータ β を求める問題で、これはあらゆる回帰モデルで学習可能である。ここで本研究では、サポートベクタ回帰モデル (SVR) を用いる。学習したパラメータを用いることで、未来の企業の価値やその動向を予測することができる。

3.3 予測結果

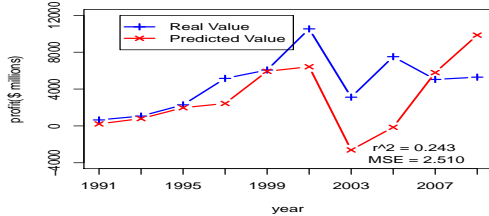
本論文では、フォーチュン 500 に 3 回以上ランキングされている企業を対象とし、ニューヨーク・タイムズから抽出したネットワークの期間 (1981 年から 2009 年まで) に合わせて、それらの企業のネットワークと企業価値 (毎年の利益値) を分析する。

まず、異なる業界から次のようなフォーチュン企業を 20 社を選ぶ: I.B.M., Intel, Microsoft, GM, HP, Honda, Nissan, AT&T, Wal-Mart, Yahoo!, Nike, Dell, Starbucks, Chase, PepsiCo, Cisco, FedEx, Gap, AEP, Sun. これらの企業は、フォーチュン 500 のリストによく現れる企業であるので、その変化する企業の価値と変化するネットワーク構造、両方を計量化することができる。ここで、SVR 学習機 (RBF カーネルを用いる) を用いて、各過去 5 年間ごとのデータを用いてモデルを学習し、次の年の利益を予測することを繰り返す。予測した年ごとの実際の利益値と予測結果を図 2 にプロットする。実際の利益値と予測値との相関と誤差は、相関係数の二乗 (r^2) と平均二乗誤差 (MSE) を用いて計算する。図で分かるように、予測のグラフは利益値の動向をほぼ攫んでいることが分かる ($r^2 = 0.440$ と $MSE = 0.437$)。ここで 1995 年の予測がかなり外れているが、その可能な理由は、これらの企業の利益は増えたが、ネットワークはその前の年の影響を受け続けていたことにある。

次に、個別企業 (IBM と Intel) の利益をそれぞれネットワークを用いて学習し予測してみた。個別企業の場合、学習データが少ないので、過去 10 年のデータに基づいて学習し、次の年の値を予測する。図 3 で示すように、IBM 社の場合 $r^2 = 0.888$ と $MSE = 0.108$ 、Intel 社の場合 $r^2 = 0.888$ と $MSE = 0.108$ であり、提案のモデルは、これらの企業の価値とその変化をほぼ予測することができた。



(a) IBM profit prediction.



(b) Intel profit prediction.

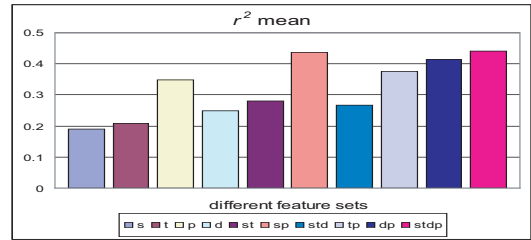
図 3: Profit Prediction for *IBM* and *Intel*.

提案したネットワークの特徴量（すなわち，ネットワーク効果）をタイプごとに評価するために，複数の特徴量の組み合わせを作って，上記の 20 企業の平均利益を予測してみる．ここで，現時点ネットワークからの特徴量 \mathbf{F}_x^t を“s”で，過去のネットワークからの特徴量 \mathbf{F}_x^H を“t”で，ネットワーク変化における特徴量 $\Delta \mathbf{F}_x^H$ を“d”で，そして企業の過去の属性値ネットワーク属性 \mathbf{P}_x^H を“p”で表す．そして，複数タイプの特徴量の組み合わせも作る．例えば，“sp”は，現時点ネットワーク効果と企業の属性値を特徴量とするし，“stdp”は，これらの 4 種類の特徴量をすべて組み合わせたものである．そして SVR (RBF カーネルを用いる) 学習機を使って，過去 5 年間のネットワークからパラメータを学習し，次の年の企業の利益値を予測した．図 4 で示すように，属性の特徴量 “p” だけを使った場合 ($r^2=0.383$, $MSE=0.287$) は，“s”，“t”，“d” のネットワークから特徴量をそれぞれ単独で使った場合より，予測の性能が高い．すなわち，相関が高くエラーが小さい．ただ，属性とネットワークの特徴量を組み合わせること（例えば，“sp”，“tp”，“dp”，“stdp”）で，予測の結果が更に高くなる事が分かる．特に全ての特徴量 “stdp” を用いた場合， $r^2=0.512$ と $MSE=0.363$ になり，この結果は，ネットワークだけと属性データだけを用いた場合よりそれぞれ 150%，34% ずつ増加することになる．

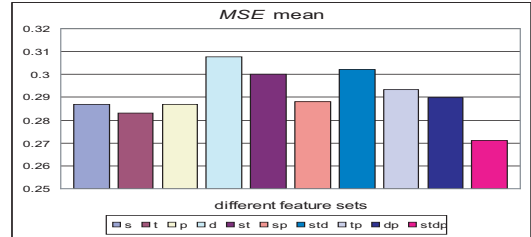
適切な *window size* と *delta size* を調べるために，これらの値をそれぞれ 1 と 3 にしてみた．その結果，両方とも 1 の方が高く，これは過去ネットワークの効果として，1 年前のネットワークを用いること，ネットワーク変化の効果としても 1 年前との変化を記録することで十分であることを示す．

4. まとめ

本研究では，企業の価値を予測する新しい分析方法を提案している．まず，公開されてニュース記事から，企業同士のインパクト関係に注目し，企業の動的ネットワークを抽出する手法を提案した．対象企業が他の企業から受けるインパクトの強さは，2つの企業名が共起する文書と文の数，そしてこれらの文書と文の重要度を計算することで計算する．抽出されたネットワークは，重み付き有向ネットワークである．次に，抽出されたネットワークから，対象企業のネットワーク効果を調べた．現在ネットワークからの効果，過去ネットワークからの効果，



(a) Profit (r^2 correlation).



(b) Profit (MSE error).

図 4: Mean profit prediction for 20 Fortune companies using different feature sets. s, structural features; t, temporal features; d, delta-change in temporal features; p, financial profiles.

そしてネットワークの変化による効果を抽出する．そして，これらのネットワーク効果の特徴量とし，SVR 学習機を使って，企業価値（今回は属性値）を学習し，未来の価値を予測した．

提案のモデルは，企業グループや個別企業の価値の変化をうまく追跡することができた．さらに，ネットワークの効果と属性データを組み合わせることで，単独のネットワーク効果，または単独の属性データを使つての予測を，それぞれ 150% と 34% ずつ上回った．さらに，*window size* と *delta size* のパラメータを調べることで，過去 1 年までのネットワーク情報とその変化を予測に用いることで十分である事が分かる．言い換えると，企業の関係やネットワーク構造は，1 年後の企業の利益に影響を及ぼしている事が分かった．

参考文献

[Wasserman 94] Wasserman, S.: Social network analysis. methods and applications, Cambridge University Press (1994).
 [Uzzi 97] Uzzi, B.: Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness, Administrative Science Quarterly, Vol.42, pp.35–67 (1997).
 [McCulloh and Carley 09] McCulloh, I. A. and Carley, K. M.: Longitudinal Dynamic Network Analysis, CASOS technical report, Center for the Computational Analysis of Social and Organizational Systems (2009).
 [Doreian and Stokman 97] Doreian, P. and Stokman, F.: Evolution of Social Networks, Amsterdam etc.: Gordon and Breach (1997).
 [Bernstein et al. 02] Bernstein, A. and Clearwater, S. and Hill, S. and Perlich, C. and Provost, F.: Discovering Knowledge from Relational Data Extracted from Business News, SIGKDD-2002 Workshop on Multi-Relational Data Mining (2002).
 [Karamon et al.] Karamon, J. and Matsuo, Y. and Ishizuka, M.: Generating Useful Network-based Features for Analyzing Social Networks, Twenty-Third Conference on Artificial Intelligence (AAAI-08), pp.1162–1168 (2008)