

ユーザー間実距離を用いたSNSユーザー評価手法とパラメータ評価

Evaluation and Development reputation network for SNS user evaluation using realistic distance

大塚 孝信^{1*} 吉村 卓也³ 伊藤 孝行^{1,2}
Takanobu Otsuka¹ Takuya Yoshimura³ Takayuki Ito^{1,2}

¹ 名古屋工業大学 グリーン・コンピューティング研究所

¹ Center for Green Computing, Nagoya Institute of Technology

² 名古屋工業大学大学院 産業戦略工学専攻

² Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology

³ 名古屋工業大学大学院 情報工学専攻

³ School of Techno-Business Administration, Graduate School of Engineering,
Nagoya Institute of Technology

Abstract: In recent years, SNS services such as Facebook, Google+, and Twitter are becoming very popular. In such services, many sources of information are posted and shared, although user rankings are hardly considered. In this paper, we consider an evaluation technique for web pages, such as HITS and PageRank, for SNS user evaluation applications and propose an algorithm using a user's real distance. We consider various parameters, including user distance, favorites, and the numbers of friends in SNSs in our evaluation technique. We propose a new reputation network to measure the reliability of SNS information.

1 はじめに

近年, Facebook に代表される SNS(Social networking Site) サービスのユーザー数が大幅に増加している。それに伴いユーザーの個人情報や悪意のあるアプリケーションや誤った情報が拡散するといった問題が増えている。特に震災時には嘘の情報やデマ等が善意のユーザーにより Twitter 上で拡散され、本当に必要な情報が埋もれてしまったといった事例も挙げられる。更に, Facebook ではアプリケーションの利用が盛んであるが, 性格診断アプリや占いといった, 一般的に多く利用されるアプリケーションを装い, ユーザーの個人情報や友人の情報を不正に取得するとともに, アプリケーションがユーザーになりすましてスパムメッセージを不特定多数に送信するといった悪質な事例が多く存在するようになってきている。最近ではユーザーの投稿数やコメント数といった SNS のアクティビティをリンク構造と考え, ユーザー評価を行うサービスも存在する

[1][2]。しかし, リンク構造は SEO 業者 (Search Engine Optimize) に代表されるような手法で偽装可能であり, リンク構造のみではユーザー評価に対する信頼性を確保するのは困難である。他にもユーザーの所属コミュニティやメッセージのやりとりによってユーザー同士の信頼度を計る研究も為されているが, パラメータをリンク構造に置き換えているものが多く, 完全ではない。本研究では情報の偽装を防ぐことを目的とし, 位置情報を利用することでユーザー間の距離をパラメータとして付加し, 友人数などの SNS 特有の各パラメータに重み付けをすることで SNS におけるユーザー評価のアルゴリズムを構築するとともに筆者の Facebook データを用いて評価実験を行った。本稿の構成を以下に示す。まず, 2 章で本研究と関連する先行研究を紹介し, 本研究の位置づけを示す。そして, 3 章で提案するアルゴリズムについて述べる。その後, 4 章において評価実験の結果と得られた知見について示す。そして最後に, 5 章で本稿のまとめと今後の課題を示す。

*連絡先: 名古屋工業大学 伊藤孝行研究室
〒466-8555 愛知県名古屋市昭和区御器所町
E-mail: otsuka.takanobu@nitech.ac.jp

2 関連研究

2.1 SNS のユーザー評価に関する研究

web のコミュニケーションは年々増加しており Facebook や Google+ に代表される SNS サービスで活発なコミュニケーションが行われている。しかし、さまざまな意見が書き込まれるものの、有用な情報のみを見つけることが困難である。また、偽の情報が拡散したり、悪意のあるアプリケーションを実行させるといった問題が挙げられている。オンラインオークションや web ページの評価をするために数多くの研究がなされている。また、SNS における「ソーシャルな強さ」を計る研究も数多くある。オンラインオークションでは web ページ評価手法である HITS を応用した ANT (Auction Network Trust) という研究があり、web ページのリンク構造をユーザーの取引情報に当てはめ、信頼度の高いユーザーをランキングすることを目的としている。[4] この研究は SNS サービスにも応用でき、取引ではなくユーザー間のコメントや友人関係に当てはめることでユーザー評価ができると考えている。しかし、コメント数や友人数は単純なスクリプトにより簡単に偽装が可能であることから“偽装しづらい”パラメータを挿入する必要があると考えている。また、WEB ページの評価手法には PageRank や HITS が多く用いられている。HITS は web ページへの authorities, hubs のリンク構造によりページの固有ベクトルを求め、その値によりページ評価を行っている。Pagerank は多くの良質なページからリンクされているページは良質であるという考え方をを用いている。この考え方は論文評価のシステムが発端であり、多くの良質な論文から引用される論文は良い論文であるという考え方を web ページに応用したものである。これらの考え方を SNS のユーザー評価に当てはめた場合、HITS はコメント回数や友人関係などのリンク構造で表すことが出来る。PageRank の場合は SNS における「良質なユーザー」をどう決定するのかという問題が残る。PageRank の計算式を読み解くと良質なページの定義は多くのページにリンクされているかどうかを主なパラメータとしているため、先ほど述べたように単純なスクリプトによりリンク構造の偽装に遭遇しやすいと考える。例えば、リンク構造を利用した口コミ評価を行う手法 [3] などもある。また、web ページとは異なり SNS 特有の友人同士のアクティビティを利用しソーシャルな繋がりを重視した研究も数多く為されている。SNS サービスにはコミュニティという概念があり、同じ学校や同じ職場、同じクラブ活動など現実世界のコミュニティと同じものや、現実世界とは関係がない趣味、嗜好のコミュニティといったものがある。これらソーシャルな繋がりがどのような要素によって決定しているかを研究したのものがある

この研究によるとソーシャルな繋がりはユーザー同士の親密が最も不覚、それらは訪問回数、友人数、友人間のメッセージのやりとりにより決定されるとあり [8]、必ずしも同一コミュニティにいるから親密とは限らない。しかし、活発な情報交換や訪問回数では友人同士のソーシャルな強さ (Social Strength) は計ることが出来るが、友人以外の有益な情報は得にくいと考える。また、本研究に類似した内容でユーザー間の距離をひとつのパラメータとして考えた研究もある [15][11] この研究はユーザーをノードとして捉え、ネットワーク・トポロジーでのパス長を距離と考えたものである。例えば友人の友人からの情報をリシェアした場合は友人の情報をシェアするよりも有益という考え方である。パス長という考え方は新しいが web ページなどのリンク構造による評価では SNS ならではの考え方でユーザー評価をできないか考えた。更に、友人の友人は友人であるという考え方にに基づき、SNS のグループ構造を可視化した研究 [16]、や VCG ネットワークを用いて信頼度を測る研究 [12] もある。これらの考え方を SNS サービスに適用した場合、リンク構造のみに評価を頼ることになるため評価の詐称がしやすいとも言える。リンク構造を悪用し、web ページの検索順位を上げるといった方法は SEO (Search Engine Optimization) 会社により多く行われている。これにより、アフィリエイトを目的とした web サイトのような内容もないページがランキング上位に来ることでユーザーにとって必要な情報が手に入りにくくなる。よって SNS の評価手法にはユーザー間の実距離を用いることで従来とは違う評価手法を提案する。特にスマートデバイスが普及してきた現在にとってはジオロケーション情報は容易に取得できる。ジオロケーションは端末側を Hack しない限りは偽装が困難であることからリンク構造と比較して高い信頼度を持つ。本研究ではユーザーのジオロケーションを用いてユーザー間の実距離を用いた評価手法を提案する。

3 ユーザ間距離を用いた評判ネットワークの提案

3.1 ユーザ間現実的距離の概念

ユーザー間の現実的距離とは SNS サイトのプロフィールに投稿された居住地や投稿情報に付与されたジオタグ等により情報をやりとりしたユーザー間の現実距離を km 単位で算出することを指す。ユーザー間距離にはユーザーの居住地同士の現実距離とユーザーの投稿した情報同士の現実距離の 2 つのパターンが存在すると考える。本アルゴリズムでは HITS や Pagerank などの単純なユーザー関係に加え、ユーザー間の現実的な距離

(Distance) を考慮することとしている。これは単純にユーザ間の現実的距離が小さい場合は現実世界でユーザ同士が顔なじみである可能性が高いと仮定しているため、通常の友人関係での情報のやり取り同様に重要ではない情報をシェアする事が多く想定される。対してユーザ間の現実的距離が離れている場合でも情報をシェアし合う仲と仮定し、現実的距離が近い場合に比べ有益な情報が多く存在していると仮定しているためである。すなわち、SNS 上において同僚や同級生同士の会話のような現実のコミュニケーションの延長でのやりとりと比較し、ユーザー間の実距離が離れていてもシェアされる情報の価値が高いと仮定している。これにより、従来リンク/被リンクのみの単純な順位付けであったものをユーザ間の現実的距離を考慮することによりリンクの重みを付加することができ、従来手法と比較した場合にユーザ評価をより正確に行うことが可能であると考えられる。

3.2 PageRank を応用したアルゴリズムの提案

Google の PageRank は「多くの良質なページからリンクされているページは、やはり良質なページである」という再帰的な関係をもとに、全てのページの重要度を判定している。PageRank とは単純な総和公式、その源は学術誌の間での論文参照構造の分析にさかのぼる公式である。[10] ページ P_i の PageRank は、 $r(P_i)$ と書くが、 P_i を指している全てのページの PageRank の総和となる。ここで、 B_p は、 P_i を指すページ（バックリンク）の集合であり、 $|P_j|$ はページ P_j からの出リンクの個数である。この際、ページ P_i の入リンクとなるページの PageRank である値 $r(P_j)$ が未知であるが、反復法を用いて解決している。すなわち、最初に全てのページが同じ PageRank の値（ウェブインデックスにあるページの個数を n として、 $1/n$ ）を持つと仮定する。そこでインデックスの各ページ P_i について $r(P_i)$ を計算する。それらを繰り返し計算することにより算出することができる。この手続きはすべてのページ P_i に対して、 $r_0(P_i) = 1/n$ として開始され、PageRank の得点が最終的には安定した値に収束するものと期待され繰り返される。ここまでの PageRank の仕組みであるが、Distance-HITS と同じくウェブページのランク付けをユーザの評価とした上でユーザ間の現実的距離情報を付加する。これにより以下計算式となる。

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \left\{ \frac{r_k(P_j)}{|P_j|} + \alpha d(P_i, P_j) \right\}$$

単純に d を足すだけではなく α を挿入することにより、パラメータの設定を容易としている。パラメータ

については評価実験を含めて実施する際に最適な値を模索していく。

3.3 SNS の各パラメータと重み付け

SNS には様々な要素が存在する。Facebook におけるパラメータを以下に示す。

- 情報を他のユーザーに拡散する - シェア
- 自分の投稿した情報が他のユーザーにより拡散される - リシェア
- 自分がフォローしている友人数
- 自分がフォローされている被友人数
- 自分の投稿した情報が他のユーザーにより評価される - 被いいね! 数
- 友人の投稿した情報を自分が評価する - いいね! 数
- 友人のウォールにコメントする - コメント数
- 自分のウォールに友人がコメントする - 被コメント数

上記のように様々なパラメータが存在するが、本研究では自分の投稿した情報が他のユーザーにより拡散される行為（リシェア）をリンク構造のパラメータとして用いている。ユーザーによる投稿間の実距離については Facebook の提供する API では取得できなかったため手作業で追加している。また、友人数についてはフォローしている友人数のみではなく、フォローされている被有人数を友人数で割ることとしている。これにより、友人数が多いだけのユーザーより、被友人数が多いユーザーの方が評価が高くなるよう配慮している。特に Facebook のような実名でのコミュニケーションを重視する SNS サービスに於いて、友人数はフォローすることで増やすことが可能であるが被友人数は相手の同意がない限り増やすことができないためある。

4 評価実験

4.1 実験設定

評価実験には筆者の Facebook データを用いている。データは Facebook の提供する Graph.API を用いており、ユーザー同士の投稿のシェアの記録を取得することができる。データには 256 人の友人（ノード）と 3568 件の投稿のシェア/リシェアの情報が記録されている。本研究ではシェアした回数よりも情報がシェアされることに重きを置くこととし、計算アルゴリズムには

Distance-Pagerank を用いることとした。実験には筆者の Facebook データを用いて独自に開発した計算アプリケーションを用いてユーザー毎の固有値ベクトルをスコアとして算出している。アプリケーションは Java によって記述されており、外部アプリケーションで出力したユーザー情報を.csv 形式で取り込むことでスコア計算を行う。開発したアプリケーションは GUI インターフェースで操作可能であり、アプリケーションでは通常の Pagerank のみでのスコア、Distance-Pagerank でのスコア、被友人数/友人数を考慮に入れたスコアを計算することが可能となっている。更に Pagerank, Distance-Pagerank と被友人数/友人数の重みを 0 から 1 の範囲で調整することが可能である。計算したスコアは.csv 形式で書き出しを可能としている。

開発したアプリケーションを 1 に示す。

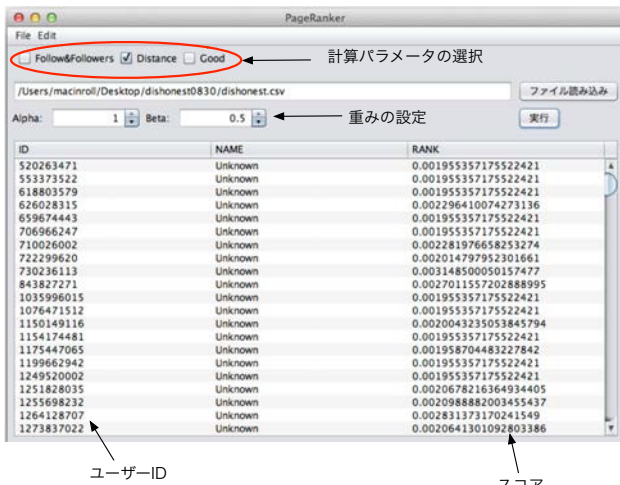


図 1: 開発したスコア計算アプリケーション

4.2 実験結果

筆者の Facebook データを用いて実際に計算した結果を示す。以下の 3 種類について計算を行っている。

1. Pagerank のみで計算
2. Distance-Pagerank での計算
3. Distance-Pagerank と被有人数/被友人数での計算

計算には開発した計算アプリケーションを使用しており、縦軸はユーザーのスコア、横軸をユーザー ID としている。散布図の作成には計算アプリケーションによって計算されたスコアを R を用いてグラフ化している。

2 に Pagerank のみでの計算結果を示す。Pagerank

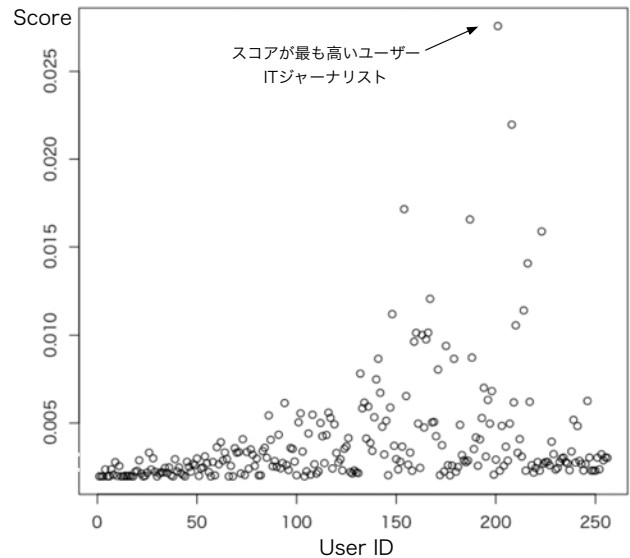


図 2: Pagerank のみでの計算結果

のみでの計算結果では投稿がシェアされることの多いアクティブユーザーの評価が高いことが分かる。筆者のネットワークで最も高いユーザーは IT ジャーナリスト、2位のユーザーは IT エバンジェリストであるため、投稿が多くシェアされていることがわかる。2位以下についてはおおまかに 2つのグループに分けられており、スコアが中間的なユーザー層、その他の多くのユーザーがスコアの低いユーザー層となっており、正規分布に類似した形となっている。

次は、Distance-Pagerank での計算結果を 3 に示す。

2 に示す Pagerank のみの結果と比較して、上位 3 位以下に変動が見られる。投稿間の実距離をパラメータとして用いることで投稿のシェアだけではなく、実距離が離れたユーザーから投稿をシェアされることでスコアが向上していることが分かる。また、下位ユーザーに関しては近くの友人や学校の同級生といった現実世界の延長としてコミュニケーションを行っているユーザーのスコアが下がっていることが分かる。距離をパラメータとして挿入することでスコアが中間的なユーザー層の順位が大きく入れ替わっており、実距離の遠いユーザーからリシェアされるユーザーのスコアが向上していることが分かる。

Distance-Pagerank と友人数での計算結果を 4 に示す。3 の Distance-Pagerank の結果と比較してあまり変化はないが、一部の低位ユーザーのスコアが変動していることが分かる。これは被友人数を友人数で割った後にパラメータとして挿入しているため、機械的に友人数を増やしているユーザー、すなわち友人数のみ極端に多く被友人数が少ないユーザーが存在しないた

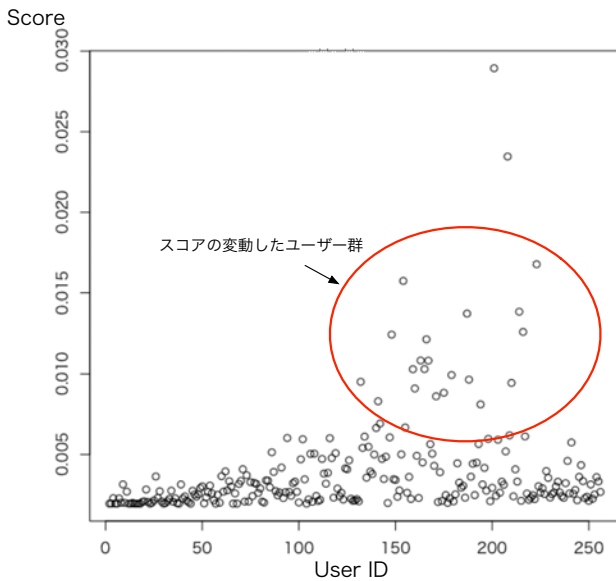


図 3: Distance-Pagerank の計算結果

め、あまり変化が見られないものとする。

4.3 考察

評価実験により、実際の Facebook データを基にし現実的な距離をユーザー評価に結びつけることでリンク構造だけではない評価手法を提案した。リンク構造のみの評価だけでなく情報間の実距離を反映することが出来た。これにより、シェアする/シェアされただけのリンク構造による評価では自動スクリプトなどにより故意に評価を上げることが出来るがユーザーのジオロケーションに紐付いた情報間の距離は偽装しにくいと考えるためより正確度の高い評価手法と言える。実際の例を見てもいたずらに情報のやり取りが多いだけではなく距離の離れているユーザーにも投稿がシェアされているという関係のほうが評価が高くなっていることがわかる。今後は SNS 特有のコメント数、いいね! 数といったパラメータに関しても実装していく。特に Facebook やでは外部サイトとの連携でいいね! 数を追加することがごく気軽に可能なため、パラメータとしては低く扱うべきだと感じている。そのため、適切な重みを付けた上で総合的に評価することを目標としている。また、現状では距離情報の取得が自動化されていないため Facebook API より自動で取得する機能を実装する必要がある。更に、計算の過程で判明した結果として少ないユーザー数で計算した場合、1つのリシェアによりスコアが大きく左右されるといった問題もある。この問題は Pagerank, HITS 共にある問題であり、全てのページ (リシェア) の総和により計算を行うため、

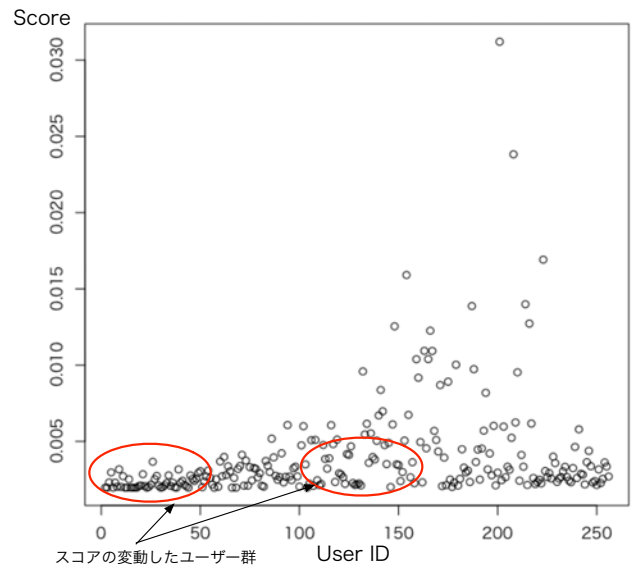


図 4: Distance-Pagerank と友人数での計算結果

リシェアの総数が少ない場合はリシェアの数が少し変動しただけで結果が大きく変わってしまう。筆者一人のデータのみではユーザー数に限界があるため、他のユーザーデータを入手することで大規模なデータを用いればこの問題を解決できると考えている。

5 まとめ

本論文では実際に Facebook のユーザー関係をネットワーク構造として捉え、既存の web ページの評価手法を用いた。これによりリンク構造だけの評価手法では故意的なスクリプトによるランキングの改ざんが可能であると考えている。しかし、ジオロケーションを外部から操作できない状態での情報間実距離を用いたランキング手法においては悪意のあるユーザーによるランキング操作が行いにくい。SNS サービスではユーザー数が増えるにつれ悪意のあるユーザーによる投稿を排除したりウイルスの埋め込まれたアプリケーションによる被害が多く報告されるようになってきている。そのため、ユーザー間距離をパラメータとして捉えることによりユーザーのランク付けの確実度を計る手法について提案した。この手法を用いることによりユーザーランクによる情報の信頼度を計ることができると考える。また、SNS サービスにあるさまざまなパラメータを組み合わせることで評価手法として確立したいと考えている。SNS サービスにはさまざまなパラメータが存在しており、SNS サービスによって異なることが多いが、各パラメータについて SNS 毎にカテゴリを設定し定量化することとしている。今後は Facebook API か

らの距離の取得を自動化することで Facebook アプリケーションとしての開発を進めるとともに他ユーザーのデータを取得することで大規模な実験を行い、評価していく。

謝辞

本研究の一部は、内閣府の先端研究助成基金助成金（最先端・次世代研究開発プログラム）により助成を受けている。

参考文献

- [1] Klout.inc, "Discover and be recognized for how you influence the world.", <http://klout.com/home>.
- [2] Overtex Group, "SNS 影響力スコアリング解析サービス Qrust.", <http://qru.st/>.
- [3] 小倉 達矢, 宍戸 開, 今藤 紀子, 山口 実靖, 浅谷 耕一, "レビューサイトにおける良質なレビューの特性とそれを考慮した評判情報の抽出に関する一考察", DEWS2008-Data Engineering Workshop, 2008.
- [4] 小林 真雄, 安藤 哲志, 伊藤 孝行, "Auction Network Trust : 電子商取引ネットワークにおけるユーザー間の関係を利用した評判メカニズム", 電子情報通信学会論文誌, Vol.J92-D, No.11.2009.
- [5] Taher H. Haveliwala, "Efficient Computation of PageRank," 1999 Stanford Technical Report.
- [6] S. Brin and L. Page, "The anatomy of a large-scale hyper textual web search engine," WWW7/Computer Networks, vol.30, no.1-7), pp. 107-117, 1998.
- [7] L. Li, Y. Shang, and W. Zhang, "Improvement of hits-based algorithms on web documents," Proceedings of WWW2002, pp. 527-535, 2002.
- [8] Eric Gilbert and Karrie Karahalios, "Predicting Tie Strength With Social Media," Proceedings of the 27th international conference on human factors in computing systems, 2009.
- [9] Josep M. Pujoi, Ramon Snguesa, and Jordi Delgado, "Extracting reputation in multi Agent Systems by Means of Social Network Topology," Proceedings of the first international joint conference on Autonomous agents and multiagent systems, pp. 467-474, 2002.
- [10] Taher H. Haveliwala, "Efficient Computation of PageRank," 1999 Stanford Technical Report.
- [11] Bloch, F. and M. O. Jackson, "The Formation of Networks with Transfers among Players", Journal of Economic Theory, 2007
- [12] Haoqi Zhang, Edith Law, Robert C. Miller, Krzysztof Z. Gajos, David C. Parkes, and Eric Horvitz, "Human Computation Tasks with Global Constraints: A Case Study," Proceedings of the ACM Conference on Human Factors in Computing, 2012.
- [13] Z. Gy. Nongyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with trust rank," Proceedings of the Thirtieth international conference on very large data bases, pp. 576-587, VLDB Endowment, 2004.
- [14] S. Pandit, D.H. Chau, S. Wang, and C. Faloutsos, "Netprobe: A fast and scalable system for fraud detection in online auction networks," Proceedings of the 16th international conference on World Wide Web (WWW'07), pp. 201-210, 2007.
- [15] Matthew O. Jackson, "SOCIAL AND ECONOMIC NETWORKS", Princeton University Press, 2008.
- [16] Paul Adams, "GROUPED: How small groups of friends are key to influence on the social web", New Riders, 2012.