

This is a post-peer-review, pre-copyedit version of an article published in the

Studies in Computational Intelligence

The final authenticated version is available online at:

https://link.springer.com/chapter/10.1007%2F978-3-319-72150-7_8

Newton's Gravitational Law for Link Prediction in Social Networks

Akanda Wahid -Ul- Ashraf, Marcin Budka, and Katarzyna Musial-Gabrys

Abstract Link prediction is an important research area in network science due to a wide range of real-world application. There are a number of link prediction methods. In the area of social networks, these methods are mostly inspired by social theory, such as having more mutual friends between two people in a social network platform entails higher probability of those two people becoming friends in the future. In this paper we take our inspiration from a different area, which is Newton's law of universal gravitation. Although this law deals with physical bodies, based on our intuition and empirical results we found that this could also work in networks, and especially in social networks. In order to apply this law, we had to endow nodes with the notion of mass and distance. While node importance could be considered as mass, the shortest path, path count, or inverse similarity (AdamicAdar, Katz score etc.) could be considered as distance. In our analysis, we have primarily used degree centrality to denote the mass of the nodes, while the lengths of shortest paths between them have been used as distances. In this study we compare the proposed link prediction approach to 7 other methods on 4 datasets from various domains. To this end, we use the ROC curves and the AUC measure to compare the methods. As the results show that our approach outperforms the other 7 methods on 2 out of the 4 datasets, we also discuss the potential reasons of the observed behaviour.

Akanda Wahid -Ul- Ashraf
Department of Computing and Informatics
Bournemouth University
Fern Barrow, Poole BH12 5BB, UK, e-mail: aashraf@bournemouth.ac.uk

Marcin Budka
Department of Computing and Informatics
Bournemouth University
Fern Barrow, Poole BH12 5BB, UK, e-mail: mbudka@bournemouth.ac.uk

Katarzyna Musial-Gabrys
School of Software
University of Technology Sydney
PO Box 123, Broadway NSW 2007, Australia, e-mail: Katarzyna.Musial-Gabrys@uts.edu.au

1 Introduction

Mining of network data has become one of the key research areas in recent years [44]. In particular, a lot of attention has been devoted to the link prediction problem in social networks [3, 29], where the goal is to discover if two nodes will become connected in the future. Link prediction is not only important for understanding and capturing the evolution of a network but also in many other real-world applications. A well-known example of link prediction is the feature called *friend suggestions* in Facebook. In Facebook, link prediction techniques are used to predict if person A is likely to become friends with person B, so that such connection suggestion can be recommended directly to both users. However, application of link prediction is much wider than in the above example. Link prediction has also been used to predict potential collaborators in business [32], in shopping recommendation [28], in patent partner recommendation [46], in predicting cell phone contact [38], in gene expression networks [7], in protein-protein interactions [2] and, in security related domain to detect suspicious communication [31].

2 Problem statement

Formally, given a snapshot of a network at time t , the link prediction problem is to identify new links that will be present in the network snapshot at time $t + 1$ [8, 17]. If a network has a set of V nodes and set of E edges at time t expressed as $G(V, E_t)$, link between a pair of vertices V_i and V_j is denoted by $L(V_i, V_j)$. For a link prediction problem, using information related to the network (topological or additional information about node's characteristics or attributes) the goal is to predict $L(V_i, V_j) \in E_{t+1}$, where $L(V_i, V_j) \notin E_t$. Based on the information exploited by link prediction methods, they can be categorised as:

1. **Unsupervised – based on topology of the network**, which only use structural information from the network, such as mutual friend count in social networks, path lengths, triad profiles etc. Some examples include methods like Katz, PageRank, and AdamicAdar [29]. These methods only require snapshot of a network topology at any given time t to make predictions for time $t + 1$, and are useful when past and non-topological information of the network is not available. They are applicable to any type of network dataset and don't require training.
2. **Supervised – based on topology of the network**, which are only applicable in networks where historical information regarding network's topology is available. For example, if snapshots of a network at $t - 1$ and t are given, then $t - 1$ is considered as historical information. Network characteristics like degree of certain nodes at time $t - 1$ can also be considered historical information. One example of such method is the Triad Transition Matrix (TTM) [19]. A wide range of machine learning approaches also fall into this category, if the topological infor-

mation such as mutual nodes, shortest distance etc. is considered as features, and link appearance is considered as class.

3. **Unsupervised – based on non-topological information of the network**, i.e. also considering non-structural information. In social networks, age, location, preferences etc. could be used as basis on which people make friends [40].
4. **Supervised – based on non-topological information of the network**, applicable to the same kind of datasets as in point 2 above. If non-structural historical information of a network is considered (with or without topological information) any binary classifier could be used to make prediction in this setting [4].

The simplest methods are the ones in the first category, widely applicable in any kind of network where only one structural snapshot is available. Despite their simplicity, these methods perform fairly well [29]. One of the biggest advantages is a wide range of scenarios where the methods are applicable, as they don't require a training phase and hence historical data, or non-topological information.

3 Proposed method

Our approach is inspired by Newton's law of universal gravitation, which states that the force exerted between two masses is proportional to the product of those masses, and inversely proportional to the squared distance between their centres [34]:

$$F = G \frac{m_1 \times m_2}{r^2}, \quad (1)$$

where F is the force between masses m_1 and m_2 , G is the gravitational constant, and r is the distance between m_1 and m_2 .

Newton derived this equation by empirical observation and inductive reasoning [12], which is an approach that we have also taken in this study. We consider mass to represent importance of a node, which can be measured in several ways. For example, different centrality measures are direct measurements of how important or central a node is in a given network. In this study we consider degree centrality of a node as equivalent to mass of that node, while the shortest path between the nodes represents their distance. Just like any physical body or object in the universe, no matter how distant the nodes are, there is a gravitational attraction between them. In the case of a social network this means that there is a non-zero probability of any two nodes forming a connection, which is intuitively correct. However, according to the small-world effect the average shortest path between any two nodes is relatively short [45]. The famous Milligram's experiment suggested that people in the USA are all linked with six links on average [43], but this number is now even lower due to people becoming more connected [5]. According to recent research, Facebook users are connected with 3.57 degrees of separation on average [15]. The small-world phenomena implies that even a short distance between two nodes could translate to low probability of two nodes having something in common or knowing each other. Equation 1 accounts for this property in an elegant manner by making

the force (which is proportional to the probability of forming a connection) inversely proportional to the squared distance.

In the context of social networks it has been established that link prediction could mainly be facilitated by utilising two kinds of information in the network: popularity and similarity [37]. Our approach supports this as a special case, where popularity acts as mass and inverse of similarity as distance. To further support this inspiration from physics in social networks, laws governing physical universe could exhibit similar properties (e.g. power-law degree distribution, high clustering coefficient) found in social networks, as it has been observed for the causal network of large-scale structure of space time in the accelerating universe [21]. Krioukov et al [21] has shown that this is mainly due to the fact that growth in social, biological, and Internet networks is asymptotically similar to the large scale causal network representing large scale structure of the spacetime. It is also possible to define distances in terms of dissimilarity in non-topological information, such as age difference, physical distance etc. A weighted sum of all these factors could be incorporated into the distance, allowing to naturally exploit non-topological information.

The above analogy leads to the following formula for calculating the score of two nodes forming a link in the future:

$$Score(V_1, V_2) = Score(V_2, V_1) = \frac{C_D(V_1) \times C_D(V_2)}{SP(V_1, V_2)^2}, \quad (2)$$

where C_D denotes the degree centrality, SP is the shortest path, and without loss of generality we have assumed $G = 1$, since in order to make a prediction, a ranked list of scores is required with their absolute values being irrelevant. Note, that if the score was to be interpreted as probability, for a particular network this could be achieved by setting the value of G as follows:

$$G = \frac{\min_{\forall (i,j), i \neq j} SP(V_i, V_j)^2}{\max_{\forall i} C_D(V_i) \times \max_{\forall j \neq i} C_D(V_j)}, \quad (3)$$

where the numerator is equal to 1, which reflects the obvious existence of a direct link between at least one pair of nodes.

It is worth noting that the proposed method in equation 2 only requires snapshot of a network at any given time. In fact, the approach is applicable even if we only have access to degree of nodes and shortest paths between them. This method falls into the *Unsupervised – based on topology of the network* category from Section 2.

Figure 1 depicts a simple social network to illustrate the intuition behind the proposed method. Application of Equation 2 produces a ranked list of scores for all pairs of nodes which are not already directly connected, given in Table 1. The following observations can be made here:

- If two people have many friends but are themselves distant (e.g. nodes 1 and 10) then the distance will make it improbable for these nodes to connect. On the other hand, someone who joins a network and have fewer friends might connect with the nearest friend who has many connections. This is why the score for nodes 10 and 16 is the same as for nodes 1 and 10, although node 16 just have one link in

the network, whereas 1 and 10 have the highest degrees. This phenomenon is intuitive in social networks as people who have just made their first connection (we assume that node 16 has joined the network relatively recently), tend to connect with people at shorter distance, who are popular (i.e. have a high degree).

- If two nodes have many connections and are close to each other yet not connected, it is very likely they have many mutual friends and will eventually connect. This is the case for nodes 2 and 10, which have the highest score of connecting according to the proposed algorithm.

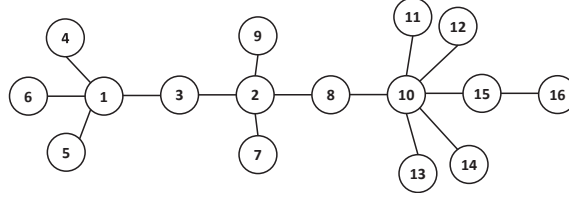


Fig. 1: Example for link prediction with a simple graph

Rank	1	2	3	4	5	6	7	8												
Score	6.0	4.0	1.5	1.5	1.33	1.0	1.0	0.89	0.89	0.67	0.67	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	
V_i	2	1	1	10	3	3	8	1	2	7	9	3	3	3	3	3	7	8	8	8
V_j	10	2	10	16	10	8	15	8	15	10	10	4	5	6	7	9	8	9	11	12

Table 1: Prediction value for a simple graph in Figure 1

In terms of computational complexity and scalability of the proposed method, if a network has a total of $|V|$ nodes and $|E|$ edges, the total number of predicted links will be the number of possible links that could exist in the graph. A fully connected undirected graph has $\frac{|V|(|V|-1)}{2}$ edges in total, so we make a prediction for at most $\frac{|V|(|V|-1)}{2} - |E|$ links. If we consider the worst case, where all the nodes are connected with with lowest amount of edges then $|E| = (|V| - 1)$. As a result, the worst case time complexity is $O(|V|^2)$, assuming that the graph is connected.

In order to calculate all shortest paths the *igraph* package uses an implementation of breadth-first search for an unweighted graph [13]. This algorithm has the worst case time complexity of $O(|V|^3)$ as all shortest paths are being calculated here [35]. However, there exist other algorithms which have time complexity of $O(|V||E|)$ for undirected graphs with integer weights [41, 42].

4 Data and Experimental Setup

For the experimental comparative evaluation of the proposed method we have selected four datasets from various domains and with different sizes, frequently used in the literature, all representing undirected graphs:

1. **hep-th**: Collaboration graph of authors of scientific papers from High Energy Physics – Theory (hep-th) section, where edges between two nodes represent a common publication. This dataset is collected from the Konect database [22, 25, 27] and has been used in the experiment of Liben-Nowell, which is a very important research work in the area of link prediction [29].
2. **contact**: Dataset representing an undirected network where edges are contacts between people using small portable wireless devices called iMotes, which were distributed among different groups of people [11, 23].
3. **hypertext**: Face-to-face contacts of ACM Hypertext 2009 conference attendees, where edges represent interactions lasting at least 20 seconds [18, 24].
4. **collegeMsg**: Private messages sent on an online social network at the University of California, Irvine [36].

Details of the datasets can be found in Table 2. To enable link prediction experiments we have split each dataset into two parts covering either equal periods of time or equal numbers of edges, where the first part is for running the algorithms and making predictions, and the second part is used for testing (ground truth). Table 2 also summarises basic properties of the split datasets. In particular, the *hep-th* dataset has been divided into part 1 containing publications between 1994 and 1996, and part 2 with publications between 1994 and 1999. The *contact*, *hypertext*, and *collegeMsg* datasets have been divided according to the number of edges. In edge stream format they are sorted according to time stamps and then split in half. Part 2 of each dataset described here contains the new edges along with the previous ones, so is effectively the full network.

	Vertices	Edges	Global clustering coefficient	Mean Degree centrality	St.Dev. Degree centrality	Average Geodesic path
hep-th 1	22398	50943	0.467	4.55	27.48	3.49
hep-th 2	22908	290484	0.333	25.36	108.48	3.22
contact 1	264	14122	0.680	106.98	248.19	2.10
contact 2	274	28244	0.566	206.16	461.77	2.42
hypertext 1	110	10409	0.396	189.25	184.56	1.78
hypertext 2	113	20818	0.495	368.46	324.43	1.66
collegeMsg 1	1260	29918	0.056	47.49	95.56	3.06
collegeMsg 2	1899	59835	0.057	63.02	131.77	3.05

Table 2: Network properties of the datasets

As it can be seen in Table 2, most of the networks have relatively high value of global clustering coefficient and short average geodesic path, which implies small-world networks [45]. Also, we could see that standard deviation of degree centrality is quite high, which implies that degree varies highly and is spread across the vertices. This variation of degree is something which affects our algorithm. With more variation of degree the proposed method will produce more spread out scores.

To benchmark our approach 7 other unsupervised link prediction methods which only use topological information were applied:

1. **Common Neighbours (CN)**, where the likelihood of two nodes A and B to develop a link depends on the number of mutual friends [33]. This method could be quantified via the following equation (Γ represents the node degree):

$$Score(A, B) = |\Gamma(A) \cap \Gamma(B)|, \quad (4)$$

2. **Jaccard Coefficient (JC)**, which is a version of Common Neighbours [16] normalised by the total number of neighbours of both nodes:

$$Score(A, B) = \frac{|\Gamma(A) \cap \Gamma(B)|}{|\Gamma(A) \cup \Gamma(B)|} \quad (5)$$

3. **Adamic-Adar Coefficient (AA)**, which is a similarity metric used in information retrieval [29] similar to the Jaccard Coefficient. In this method the likelihood of two nodes connecting depends on the number of common neighbours (e.g. mutual friends in a social network) relative to the nodes' degrees [1]. In Equation 6, C denotes rare features (in this case a rare feature is having many mutual friends while the degree of both nodes is low), common between nodes A and B :

$$Score(A, B) = \sum_{C \in \Gamma(A) \cap \Gamma(B)} \frac{1}{\log |\Gamma(C)|} \quad (6)$$

As an example, consider nodes A (with 500 friends) and B (with 300 friends) which are not yet connected and have 30 mutual friends. Now, if another two not yet connected nodes X and Y have 100 and 70 friends respectively, and also 30 mutual friends, using this formula the (X, Y) pair will get a higher score although the number of mutual friends between both pairs is the same.

4. **Preferential Attachment (PA)**, based on the social theory of 'rich get richer' implying that nodes with higher degree are more likely to get new links [6]:

$$Score(A, B) = |\Gamma(A) \cdot \Gamma(B)| \quad (7)$$

5. **Katz**, which considers the number of all the paths from node A to B [20]. The shorter paths have more weight, which is damped exponentially with path length and the β parameter (M denotes the adjacency matrix):

$$Score(A, B) = \beta M + \beta^2 M^2 + \beta^3 M^3 + \dots \quad (8)$$

β needs to be smaller than the reciprocal of the highest eigenvalue of M [26]. In our experiments $\beta = 0.001$ was used as it fits this criterion for all 4 datasets.

6. **Rooted PageRank (RPR)**, which is used by search engines to rank websites. In graph analysis it works by ranking nodes, and the rank is determined the probability of each node being reached via random walk on the graph [9]. The $Score(A, B)$ is calculated using the stationary probability distribution of B in a random walk. The random walk returns to A with the probability α at each step, moving to a random neighbour with probability $1 - \alpha$ [29]. Liben-Nowell and

Kleinberg [29] also found that setting $\alpha = 0.15$ produces best performance on most of the datasets when compared with a random predictor.

7. **Degree**, which is based on the degree of the target node used as a score [30, 31].

5 Results

All the benchmarking methods described in Section 4 produce a ranked list of edges with scores, which indicate how likely they are to be created in the future. After applying each of the methods on part 1 of the four datasets, we have used part 2 as ground truth for measuring performance. We plot the Receiver Operating Characteristic curve (ROC) and calculate the Area Under the Curve (AUC) to compare the link prediction methods. To calculate and plot the ROC curve we have used a widely used and tested opensource R package called pROC [39].

	Proposed	Adamic Adar	Common Neighbor	Degree	Jaccard Coeff.	Katz	Preferential Attachment	Rooted PageRank 0.15
hep-th	0.864	0.865*	0.864	0.737	0.852	0.901***	0.834	0.897**
contact	0.892***	0.879	0.878	0.728	0.810	0.881*	0.891**	0.798
hypertext	0.638***	0.626*	0.622	0.604	0.606	0.621*	0.638***	0.634**
collegeMsg	0.811**	0.662	0.660	0.754	0.635	0.791*	0.812***	0.791*

Table 3: AUC comparison(*** highest, ** 2nd highest, * 3rd highest)

The results have been depicted in Table 3 and Figure 2. As it can be seen the proposed approach outperforms the remaining methods for the *hypertext* and *contact* datasets, is the 2nd best for the *collegeMsg* dataset and 4th best for *hep-th*. It is interesting to note that even without considering mutual friends in these networks and only taking the degree and shortest path into consideration, we can obtain state-of-the-art prediction results. For the *hypertext* dataset all methods perform relatively poorly, with much smaller differences in terms of AUC. One important observation for the *hypertext* dataset is that Preferential Attachment and our proposed approach have the same AUC. This could be due to the average and maximum geodesic path being only 1.66 (Table 2). In this study we only utilise geodesic path as distance, and a wider range of geodesic path entails better performance. It can also be seen that our approach is predominant when compared with Preferential Attachment for *hep-th* and *contact* dataset. This shows that dividing by distance performs better than mere multiplication of degrees of predicting node couples. This difference is more visible for the *hep-th* dataset as average geodesic path of 3.49 is also the highest among all of the tested datasets.

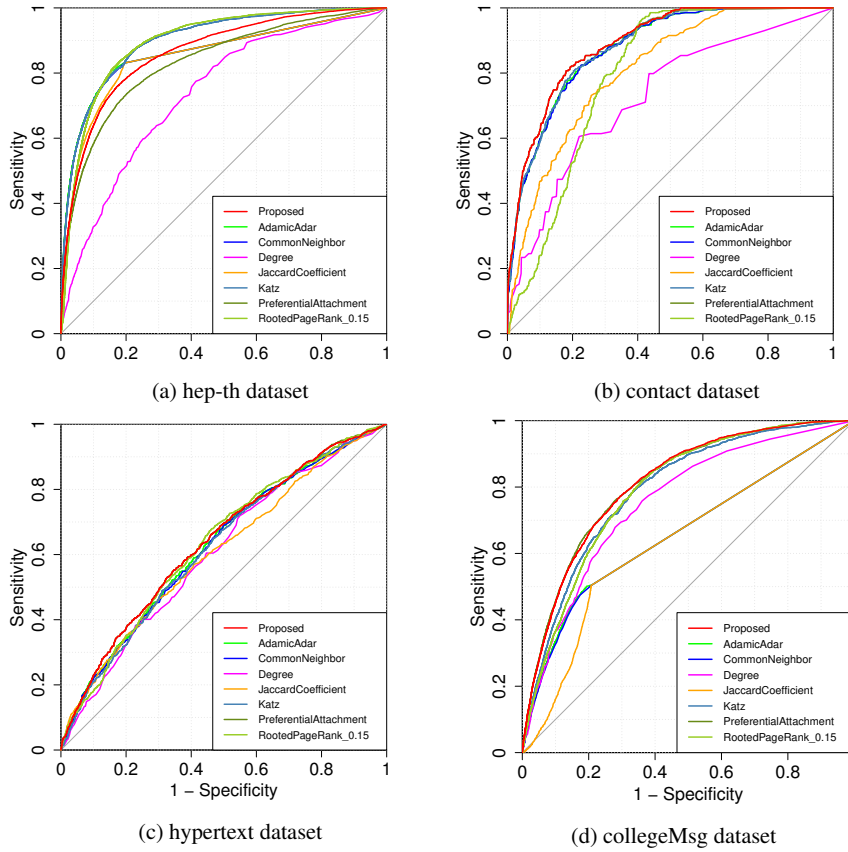


Fig. 2: ROC curves for four datasets with different link prediction methods

6 Conclusions and future work

This analysis on the four datasets shows that the proposed approach works well and could be used as a link prediction method. However, using other centrality measures and/or different definitions of distance, could result in even better performance, at least in some cases. It is important to note that we are proposing a method which combines multiplication of weights/importance divided by quadratic distance/dissimilarity, and definition of importance and distance could go beyond the topology of the network. For instance, in social networks geographic distances, age differences and, political differences between two people could be incorporated into the distances. Even relying solely on topology, scores produced by all the other link prediction methods (e.g. AdamicAdar, Katz) could potentially be used to quantify the distance. Similarly features like number of publications in a citation network or academic qualifications could quantify importance (i.e. mass), so could betweenness/closeness centrality if only topological information was available.

The work presented in this paper opens a number of potential future research avenues. The proposed method requires further validation on a larger volume of datasets in order to confirm its performance and better understand in which situations it performs well. We also acknowledge that the AUC-based evaluation has its limitations due to imbalanced classes [10, 14], and intend to also incorporate Precision and Recall in our future work. The impact of using other centrality measures such as betweenness centrality, closeness centrality or their combination, as well as various definitions of distance (e.g. potentially incorporating information on node characteristics like age or gender) also needs to be investigated. A particularly interesting question arising here is if it is possible to understand how the definitions of mass and distance affect performance in the context of a particular network structure, which would be akin to a meta-learning approach. Finally, a supervised variant of the method using historical information is an important research direction.

References

- [1] Adamic LA, Adar E (2003) Friends and neighbors on the web. *Social networks* 25(3):211–230
- [2] Airoldi EM, Blei DM, Fienberg SE, Xing EP, Jaakkola T (2006) Mixed membership stochastic block models for relational data with application to protein-protein interactions. In: *Proceedings of the international biometrics society annual meeting*, pp 1–34
- [3] Al Hasan M, Zaki MJ (2011) A survey of link prediction in social networks. In: *Social network data analytics*, Springer, pp 243–275
- [4] Al Hasan M, Chaoji V, Salem S, Zaki M (2006) Link prediction using supervised learning. In: *SDM06: workshop on link analysis, counter-terrorism and security*
- [5] Backstrom L, Boldi P, Rosa M, Ugander J, Vigna S (2012) Four degrees of separation. In: *Proceedings of the 4th Annual ACM Web Science Conference*, ACM, pp 33–42
- [6] Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications* 311(3):590–614
- [7] Barzel B, Barabási AL (2013) Network link prediction by global silencing of indirect correlations. *Nature biotechnology* 31(8):720–725
- [8] Bliss CA, Frank MR, Danforth CM, Dodds PS (2014) An evolutionary algorithm approach to link prediction in dynamic social networks. *Journal of Computational Science* 5(5):750–764
- [9] Brin S, Page L (2012) Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks* 56(18):3825–3833
- [10] Cannistraci CV, Alanis-Lobato G, Ravasi T (2013) Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* 29(13):i199–i209

- [11] Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J (2007) Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing* 6(6)
- [12] Crombie A (1957) Newton's conception of scientific method. *Physics Bulletin* 8(11):350
- [13] Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5):1–9
- [14] Davis J, Goadrich M (2006) The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd international conference on Machine learning*, ACM, pp 233–240
- [15] Edunov S, Diuk C, Filiz IO, Bhagat S, Burke M (2016) Three and a half degrees of separation. *Research at Facebook*
- [16] Gerard S, Michael JM (1983) *Introduction to modern information retrieval*. McGraw-Hill, New York
- [17] Hristova D, Noulas A, Brown C, Musolesi M, Mascolo C (2016) A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Science* 5(1):24
- [18] Isella L, Stehlé J, Barrat A, Cattuto C, Pinton JF, Van den Broeck W (2011) What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology* 271(1):166–180
- [19] Juszczyszyn K, Musial K, Budka M (2011) Link prediction based on subgraph evolution in dynamic social networks. In: *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, IEEE, pp 27–34
- [20] Katz L (1953) A new status index derived from sociometric analysis. *Psychometrika* 18(1):39–43
- [21] Krioukov D, Kitsak M, Sinkovits RS, Rideout D, Meyer D, Boguñá M (2012) Network cosmology. *Scientific reports* 2
- [22] Kunegis J (2013) arxiv hep-th network dataset konect. <http://konect.uni-koblenz.de/networks/ca-cit-HepTh>, accessed: April 2017
- [23] Kunegis J (2013) Huggle network dataset konect. <http://konect.uni-koblenz.de/networks/contact>, accessed: April 2017
- [24] Kunegis J (2013) Hypertext 2009 network dataset konect. <http://konect.uni-koblenz.de/networks/contact>, accessed: April 2017
- [25] Kunegis J (2013) Konect: the koblenz network collection. In: *Proceedings of the 22nd International Conference on World Wide Web*, ACM, pp 1343–1350
- [26] Landherr A, Friedl B, Heidemann J (2010) A critical review of centrality measures in social networks. *Business & Information Systems Engineering* 2(6):371–385
- [27] Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1(1):2
- [28] Li J, Zhang L, Meng F, Li F (2014) Recommendation algorithm based on link prediction and domain knowledge in retail transactions. *Procedia Computer Science* 31:875–881

- [29] Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology* 58(7):1019–1031
- [30] Lichtenwalter RN, Chawla NV (2011) Lpmed: Link prediction made easy. *Journal of Machine Learning Research* 12(Aug):2489–2492
- [31] Lichtenwalter RN, Lussier JT, Chawla NV (2010) New perspectives and methods in link prediction. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 243–252
- [32] Mori J, Kajikawa Y, Kashima H, Sakata I (2012) Machine learning approach for finding business partners and building reciprocal relationships. *Expert Systems with Applications* 39(12):10,402–10,407
- [33] Newman ME (2001) Clustering and preferential attachment in growing networks. *Physical review E* 64(2):025,102
- [34] Newton I (1987) *Philosophiæ naturalis principia mathematica* (mathematical principles of natural philosophy). London (1687)
- [35] Pandey VS, Kumar R, Singh P (2010) An optimized all pair shortest paths algorithm. *International Journal of Computer Applications* 2(3):0975–8887
- [36] Panzarasa P, Opsahl T, Carley KM (2009) Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the Association for Information Science and Technology* 60(5):911–932
- [37] Papadopoulos F, Kitsak M, Serrano MÁ, Boguñá M, Krioukov D (2012) Popularity versus similarity in growing networks. *Nature* 489:537–540
- [38] Raeder T, Lizardo O, Hachen D, Chawla NV (2011) Predictors of short-term decay of cell phone contacts in a large scale communication network. *Social Networks* 33(4):245–257
- [39] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M (2011) proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC bioinformatics* 12(1):77
- [40] Tan R, Gu J, Chen P, Zhong Z (2013) Link prediction using protected location history. In: *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, IEEE, pp 795–798
- [41] Thorup M (1999) Undirected single-source shortest paths with positive integer weights in linear time. *Journal of the ACM (JACM)* 46(3):362–394
- [42] Thorup M (2000) Floats, integers, and single source shortest paths. *Journal of algorithms* 35(2):189–201
- [43] Travers J, Milgram S (1967) The small world problem. *Psychology Today* 1:61–67
- [44] Wang P, Xu B, Wu Y, Zhou X (2015) Link prediction in social networks: the state-of-the-art. *Science China Information Sciences* 58(1):1–38
- [45] Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *nature* 393(6684):440–442
- [46] Wu S, Sun J, Tang J (2013) Patent partner recommendation in enterprise social networks. In: *Proceedings of the sixth ACM international conference on Web search and data mining*, ACM, pp 43–52