Track: 9th International Workshop on Modeling Social Media (MSM 2018)
Applying Machine Learning and AI for Modeling Social Media

WWW 2018, April 23-27, 2018, Lyon, France

Structural Novelty and Diversity in Link Prediction

Javier Sanz-Cruzado
Universidad Autónoma de Madrid
javier.sanz-cruzado@uam.es

Sofía M. Pepa
GMV Information Technologies and Services
sofiamarinapepa@gmail.com

Pablo Castells
Universidad Autónoma de Madrid
pablo.castells@uam.es

ABSTRACT

Link prediction has mainly been addressed as an accuracy-targeting problem in the social networks field. We discuss different perspectives on the problem considering other dimensions and effects that the link prediction methods may have on the social network where they are applied. Specifically, we consider the structural effects the prediction can have if the predicted links are added to the network. We consider further utility dimensions beyond prediction accuracy, namely novelty and diversity. We discuss the adaptation, for this purpose, of specific network, novelty and diversity metrics from social network analysis, recommender systems, and information retrieval.

1 INTRODUCTION

Link prediction can be considered today one of the classic areas in social network analysis research and development [1,14,23,24]. The problem consists in finding links in a social network that have not been observed or formed yet, but may do so in the future, or may simply be useful to add. A paradigmatic application for the problem is recommending contacts in online social networks [16,17,32], a feature most popular platforms, such as Facebook, Twitter or LinkedIn, nowadays provide [13,15].

A link prediction method can be evaluated in different ways, depending on the specific nuances in how the problem is stated. If seen as a classification task, the methods can be evaluated in terms of the predictive accuracy by usual metrics such as AUC, contingency tables, etc. [23]. If stated as a recommendation problem, information retrieval metrics can be used, such as precision, recall, etc. [16]. Yet as far as we are aware, most of the evaluation approaches to date, and therefore the solutions developed targeting them, seek to optimize microscopic perspectives: the number of correctly classified links (true positives), the accuracy or the benefit the recommendation brings to each target user in isolation, to be just averaged over the network.

In this paper we contend that we may want to consider structural and global impact of link prediction, in addition to the plain accuracy and local effects, when assessing a link prediction algorithm. Contact recommendation functionalities nowadays account for an increasing fraction of the online social network growth. Link prediction therefore represents an opportunity to favor trends towards desirable global properties in the evolution of a network, beyond (and compatible with) the short-term micro level value to be procured by the recommended links.

In order to explore this perspective, the social network analysis field provides a profuse array of concepts, metrics and analytic methods to assess the properties of the effect of link recommendation on a social network. We hence explore using such notions and measures to define new evaluation metrics for link prediction. Moreover, the recommender systems field has developed over recent years a clear awareness that accuracy alone is just a rather partial view on the value of recommendation: novelty and diversity perspectives can be as important –at both the macro and micro levels. We therefore likewise consider the adaptation of outcomes from that area. We find that, at more than one level, the global network analysis dimension of edge prediction links to similar principles as lay beneath the novelty and diversity perspectives.

2 RELATED WORK

Incipient research has considered the effects of contact recommendation algorithms on global properties of the network. We can distinguish two main perspectives in this scope. The first one focuses on the measurement of the effects of recommender systems on the structure of networks. The effect on metrics such as the clustering coefficient [8,19,31], the number of connected components [19] or the degree distribution [8] have been analyzed. The second line considers influencing the network growth towards some desired properties. In particular, Parotsidis [29] seeks to minimize the expected path length between the target user and the rest of the network; and Wu et al. [35] seek to maximize the modularity of the network. In this paper, we aim to broaden the perspective undertaken in such initial research, towards a wider range of network metrics, and dimensions beyond accuracy, such as novelty and diversity.

3 NOTATION

We shall use the following notation in the rest of the paper. We denote as $\mathcal{G} = \langle \mathcal{U}, E \rangle$ the graph structure of a social network, where $\mathcal{U}$ represents the set of people in the network, and $E \subset \mathcal{U}_*^2$ represents the relations between people, $\mathcal{U}_*^2$ being a shortcut to denote the set of all different user pairs. For a person $u \in \mathcal{U}$ we

Track: 9th International Workshop on Modeling Social Media (MSM 2018)
Applying Machine Learning and AI for Modeling Social Media

WWW 2018, April 23-27, 2018, Lyon, France

denote by $\Gamma(u)$ the set of people to which u is connected. In directed networks, we shall differentiate between the incoming and outgoing neighborhoods $\Gamma_{in}(u)$ and $\Gamma_{out}(u)$ respectively.

The link prediction problem can be stated as identifying the subset of links $\hat{E} \subset (\mathcal{U}_*^2 - E)$ that are not observed but present in the network, or will form in the future, or would be useful to add −whatever the variant of the problem is. From a recommendation perspective, we shall denote by $\hat{\Gamma}_{out}(u)$ the set of people involved in the predicted arcs going out from u, i.e. $\hat{\Gamma}_{out}(u) = \{v \in \mathcal{U} | (u,v) \in \hat{E}\}$. And we shall refer to the graph including only the recommended links as $\hat{\mathcal{G}} = \langle \mathcal{U}, \hat{E} \rangle$.

4 SOCIAL NETWORK ANALYSIS

One way to assess the effect of a prediction algorithm on the network is to consider the extension of the network $\mathcal{G}' = \langle \mathcal{U}, E' \rangle$, with $E' = E \cup \hat{E}$, by a certain subset $\hat{E}$ of predicted links (for instance, the union of the top k predicted outgoing links in the ranking for each person u), as if the party the prediction is delivered to (e.g. the users of an online social network) accepted all the links in $\hat{E}$. Hence, any network metric applied to $\mathcal{G}'$ can be taken as a metric on the prediction method. We summarize here some classical metrics we find of potential interest for the perspective under discussion.

4.1 Distance-Based Metrics

An effect of recommendation, inasmuch as it increases the network density, is a general reduction of distances in the augmented graph. We may hence consider the metrics that measure this effect in different ways. We define reciprocal versions of the metrics when appropriate, in such a way that high values "are good" (in the sense that they reflect a possibly desired property).

- **Average reciprocal shortest path length:**

$$\text{ARSL}(\mathcal{G}') = \frac{1}{|\mathcal{U}|(|\mathcal{U}| - 1)} \sum_{u,v \in \mathcal{U}} \frac{1}{\delta'(u,v)}$$

  where $\delta'(u,v)$ denotes the shortest-path distance between u and v in the extended network $\mathcal{G}'$.

- **Reciprocal diameter:** $\text{RD}(\mathcal{G}') = 1/\max_{u \in \mathcal{U}} ecc(u)$

  where the eccentricity $ecc(u) = \max_{v \in \mathcal{U}: \delta'(u,v) < \infty} \delta'(u,v)$ of a node u is defined as the distance to the farthest accessible node from u in the network [9].

- **Reciprocal average eccentricity:**

$$\text{RAE}(\mathcal{G}') = |\mathcal{U}| \Big/ \sum_{u \in \mathcal{U}} ecc(u)$$

- **Mean prediction distance:**

$$\text{MPD}(\hat{E}|\mathcal{G}) = |\hat{E}| \Big/ \sum_{(u,v) \in \hat{E}} \frac{1}{\delta(u,v)} - 2$$

  where $\delta(u,v)$ denotes the shortest distance in the original network $\mathcal{G}$ (before prediction). MPD computes the harmonic mean of the prediction distances, subtracting 2 to set the metric minimum value at 0.

Distance shortening is likely a desirable effect in most cases, as it makes people easier to reach from each other through a smaller number of hops through common acquaintances. The dif-ferent ways to average the distances provides nuances in the perspective with which distance is accounted for. ARSL, RD and RAE range in (0,1], and MPD takes values in $[0, \infty]$. ARSL and MPD are defined in such a way that so-called global bridges [12], if any, between previously separate connected components are rewarded as the ideal case, whereas RD and RAE just ignore such improvements and only consider distances within components. MPD measures in a quite direct way how the predicted links bring people far from their usual social environment, which can be seen as a measure of novelty from a contact recommendation perspective.

4.2 Structural Diversity

Notions of structural diversity have been a profuse object of study in the field of complex networks [19]. From the simplest perspective, the degree distribution can be seen as a primary sign of connective diversity: a very skewed distribution reflects a concentration of links around a few highly connected people, whereas in a flatter distribution each person gets a more distinctive social circle of her own. The "flatness" of the degree distribution can be summarized by a single number using the Gini index [11], which we can reverse into the **degree Gini complement**:

$$\text{DGC}(\mathcal{G}') = 1 - \frac{1}{|\mathcal{U}| - 1} \sum_{i=1}^{|\mathcal{U}|} (2i - |\mathcal{U}| - 1) \frac{|\Gamma'(u_i)|}{|E'|}$$

where people u_i in the above definition are ordered by non-decreasing degree $|\Gamma'(u_i)|$, and Γ' represents neighborhoods in the extended graph $\mathcal{G}'$. We take the complement of the Gini index, in such a way high values indicate that the edges are evenly distributed. In directed networks it makes also sense to compute indegree and outdegree versions IDGC and ODGC.

Richer notions of structural diversity have been studied, related to the concept of weak tie. Granovetter hypothesized that such links provide more novel information that strong ties [10,12]. Measures of link strength can be defined based just on topological properties in the network, and can be related in some way or other to the sociological notion of weak tie [10,12]. Such measures are typically related to notions of redundancy: a tie is weak inasmuch as it is not redundant to other links around it; it carries a somehow exclusive (and hence valuable) connection between specific people or regions of the network. Such measures can be broadly divided in two categories: global and local.

4.1.1 Local notions

Granovetter provided the first local notion of weak link under the name of *local bridge* [12]: a link between people who do not have any common neighbors. We find this definition is rather binary and restrictive, and results in a quite coarse metric. The so-called link embeddedness provides a finer and more informative metric, which measures the relative overlap of the neighborhoods of its endpoints [36] as an indication of link strength:

2

Track: 9th International Workshop on Modeling Social Media (MSM 2018)
Applying Machine Learning and AI for Modeling Social Media

WWW 2018, April 23-27, 2018, Lyon, France

$$\text{Embeddedness}(u,v|\mathcal{G}') = \frac{|\Gamma'_{\text{out}}(u) \cap \Gamma'_{\text{in}}(v)|}{|\Gamma'_{\text{out}}(u) \cup \Gamma'_{\text{in}}(v)|}$$

We may assess the degree to which a link prediction method suggests weak ties by measuring the **average edge weakness** of the suggested links as the complement of embeddedness:

$$\text{AEW}(\hat{E}|\mathcal{G}') = \frac{1}{|\hat{E}|} \sum_{(u,v)\in\hat{E}} \left(1 - \text{Embeddedness}(u,v|\mathcal{G}')\right)$$

The metric ranges in $[0,1]$ in such a way that the higher the weakness, the higher the structural diversity brought by link prediction.

Another classical means to assess the degree of connection redundancy is the clustering coefficient. The global clustering coefficient of a network can be measured as the ratio of triangles in the network over the number of triads (paths of length two). Again, we take the **clustering coefficient complement**, to get the metric values properly aligned with a notion of diversity:

$$\text{CCC}(\mathcal{G}') = 1 - \frac{|\{(u,v,w)|(u,v),(v,w),(u,w)\in E'\}|}{|\{(u,v,w)|(u,v),(v,w)\in E'\}|}$$

A link prediction method brings diversity to the network to the extent that this metric gets a high value.

4.1.2 Global notions

Along with the concept of local bridge, Granovetter proposed a global notion of bridge: a unique link between connected components. Again, we find this definition very restrictive in common social networks, which typically display a giant connected component [25]. We hence consider a relaxed definition based on work by De Meo et al. [10]: links between communities are considered weak, and links inside communities are considered strong. Inspired by this notion, we can consider different metrics that assess the presence of such links in the network. If communities are not a given, such metrics depend on a community detection algorithm [3,7,30], but the analysis can be stable over such algorithms [10].

A classical measure of the presence of inter-community links is the so-called modularity [7,26]. Given a partition of the network into a set of communities $\mathcal{C}$, modularity compares the number of edges inside communities (strong links) to the expected number of strong ties we would find if the edges were placed at random:

$$\text{Mod}(\mathcal{G}'|\mathcal{C}) = \frac{\sum_{u,v\in\mathcal{U}}(A_{uv} - |\Gamma'_{\text{out}}(u)||\Gamma'_{\text{in}}(v)|/|E'|)[c(u) = c(v)]}{|E'| - \sum_{u,v\in\mathcal{U}}|\Gamma'_{\text{out}}(u)||\Gamma'_{\text{in}}(v)|/|E'|\,[c(u) = c(v)]}$$

where $c(u) \in \mathcal{C}$ denotes the community that u belongs to, A_{uv} is equal to 1 if there is a link between users u and v, and 0 otherwise, and $[\,\cdot\,]$ is the indicator function, which is equal to 1 iff the predicate inside the brackets is true. Once again, since low modularity indicates high diversity, we linearly reorient the values into a **modularity complement** metric ranging in $[0,1]$:

$$\text{MC}(\mathcal{G}'|\mathcal{C}) = \left(1 - \text{Mod}(\mathcal{G}'|\mathcal{C})\right)/2$$

MC provides a measure of the abundance of weak ties across communities, but it does not provide information about the distribution of the weak links over the communities. We may hence

want to consider a finer metric that assesses how balanced is the weak link distribution. For this purpose we propose the **inter-community Gini complement**, which counts the links between each pair of different communities, and computes the (complement of the) Gini coefficient [11] of the distribution:

$$\text{ICGC}(\mathcal{G}'|\mathcal{C}) = 1 - \frac{1}{M-1}\sum_{i=1}^{M}(2i - M - 1)p((c_1,c_2)_i|\mathcal{G}',\mathcal{C})$$

where $M = |\mathcal{C}|(|\mathcal{C}| - 1)$ is the number of pairs of (different) communities in the partition (M is half that value if $\mathcal{G}'$ is undirected), $(c_1,c_2)_i$ is the i-th pair of communities with the smaller number of links between them, and $p((c_1,c_2)|\mathcal{G}',\mathcal{C})$ is the probability of randomly selecting a weak tie between that pair of communities:

$$p((c_1,c_2)|\mathcal{G}',\mathcal{C}) = \frac{|\{(u,v)\in E'|c(u) = c_1 \wedge c(v) = c_2\}|}{|\{(u,v)\in E'|c(u) \neq c(v)\}|}$$

This metric has the extreme value 0 when only two communities have links across them, and 1 when every two pairs of communities have the same amount of crossing links. However, it does not inform of the total number of weak ties.

5 NOVELTY AND DIVERSITY

Diversity is a rich concept that is studied in many different disciplines. Pertinent to our present focus, the information retrieval and recommender systems fields have developed notions of their own in this scope [2,4,6,33], which can be meaningful in the link prediction context as well. We hence consider their adaptation in a perspective where link prediction is seen as a contact recommendation task, targeted to the social network users.

5.1 Novelty

Novelty is a primary concern to recommender systems in most common scenarios where recommendation is tied to a purpose of discovery [4]. The most common novelty notion refers to recommending minority items in the long tail of the popularity distribution. A **long-tail novelty** metric can be formalized as the prior probability that a random person in the network was not acquainted to some of the recommended people to some other random user:

$$\text{LTN}(\hat{E}|\mathcal{G}) = \frac{1}{|\hat{E}|} \sum_{(u,v)\in\hat{E}} \left(1 - \frac{|\Gamma_{\text{in}}(v)|}{|\mathcal{U}|}\right)$$

which is inversely equivalent to the average indegree of the predicted contacts. This metric was proposed as the *expected popularity complement* in the context of recommender systems [4,33].

While LTN measures novelty from a global perspective (how novel are links to anyone), it also makes sense to consider the specific novelty from the individual viewpoint of each particular target user. So-called **unexpectedness** metrics have been proposed in the evaluation of recommender systems [4], which assess the dissimilarity between the recommended items and the prior experience of the specific target user. In our case the available records of user experience simply consist of their present contacts in $\mathcal{G}$:

3

Track: 9th International Workshop on Modeling Social Media (MSM 2018)
Applying Machine Learning and AI for Modeling Social Media

WWW 2018, April 23-27, 2018, Lyon, France

$$\text{Unexp}(\hat{E}|\mathcal{G}) = \frac{1}{|\hat{E}|} \sum_{(u,v)\in\hat{E}} \frac{1}{|\Gamma_{\text{out}}(u)|} \sum_{w\in\Gamma_{\text{out}}(u)} d(v,w)$$

where the distance measure $d(v,w)$ between users can be defined in any meaningful way for the domain at hand. Dissimilarity measures based on side-information about users are usually suitable for this purpose.

LTN and unexpectedness can measure how much the recommended contacts takes target users far from their comfort zone, hence bringing opportunities for broadening and diversifying their social experience.

5.2 Diversity

From a recommendation perspective, the diversity of a set of predicted links refers to how different are from each other the people recommended at the other end of the links. This is commonly measured by the **intra-list dissimilarity** [4], defined as the average pairwise distance between the people recommended to each target user:

$$\text{ILD}(\hat{\mathcal{G}}) = \frac{1}{|\hat{E}|} \sum_{(u,v)\in\hat{E}} \sum_{w\in\hat{\Gamma}_{\text{out}}(u)} \frac{d(v,w)}{|\hat{\Gamma}_{\text{out}}(u)|}$$

where $d(v,w)$ is a dissimilarity measure between users, and can be defined in any meaningful way –again, distance functions on user features tend to be suitably informative.

The second important diversity notion in recommendation concerns a global perspective and has often been referred to as the *diversity of sales* [4]. In our context it can be defined as how evenly are the recommendations distributed over all users. Again, the Gini index is a suitable summary metric to assess this aspect, which we use to define the **prediction Gini complement**:

$$\text{PGC}(\hat{\mathcal{G}}) = 1 - \frac{1}{|\mathcal{U}|-1} \sum_{i=1}^{|\mathcal{U}|} (2i - |\mathcal{U}| - 1) \frac{|\hat{\Gamma}_{\text{in}}(v_i)|}{|\hat{E}|}$$

where v_i represents the i-th user by non-decreasing number of times $|\hat{\Gamma}_{\text{in}}(v_i)|$ she is recommended. This metric is equal to 1 when all users are recommended equally often, and 0 when all the predicted links point to the same single user.

5.2.2 Information retrieval perspective

A related but different take on diversity has been developed in search-oriented information retrieval (IR), that considers returning diverse results considering the different possible intents or *aspects behind* an ambiguous search query [36]. Although link prediction does not involve explicit queries, it is possible to adapt this perspective by matching users in the network to queries and documents. For instance, latent communities can be considered in the role of query aspects, whereby we can adapt all the aspect-based diversity metrics from IR.

The simplest metric, *subtopic recall*, which we may rename as **community recall** in our context, counts and averages the ratio of communities covered by the people recommended to each target user [35]:

$$\text{CRecall}(\hat{\mathcal{G}}|\mathcal{C}) = \frac{1}{|\mathcal{U}||\mathcal{C}|} \sum_{u\in\mathcal{U}} \left| \bigcup_{v\in\hat{\Gamma}_{\text{out}}(u)} c(v) \right|$$

where $\mathcal{C}$ is the set of communities of the graph, $c(v)$ is the community that user v belongs to.

A more elaborate approach to aspect-based evaluation is the so-called intent-aware scheme, in which one of the most meaningful and widely used metrics is **ERR-IA** [5]. This metric weighs down the added value of correctly predicted links in the ranking (for a given user) when the community of the recommended endpoint already occurs above in the ranking:

$$\text{ERR-IA}(\hat{E}|\mathcal{G}, E_{\text{test}}, \mathcal{C}) = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \sum_{c\in\mathcal{C}} p(c|u) \, \text{ERR-IA}(u,c)$$

$$\text{ERR-IA}(u,c) = \sum_{k=1}^{|\hat{\Gamma}_{\text{out}}(u)|} \frac{1}{k} p(\text{rel}|v_k,c) \prod_{j=1}^{k-1} \left(1 - p(\text{rel}|v_j,c)\right)$$

where v_k is the user at position k in the ranking of recommended links for user u, $p(\text{rel}|v,c)$ is commonly defined as $p(\text{rel}|v,c) = 0.5 \cdot [(u,v) \in E_{\text{test}} \wedge c(v) = c]$, and E_{test} represents the set of test links (held out from the prediction algorithms) with which the accuracy of link prediction is evaluated. The probability $p(c|u)$ that a community is pertinent to a user can be estimated by the ratio of followers of u that belong to c:

$$p(c|u) = \frac{|\{v \in c(u)|(u,v) \in E \cup E_{\text{test}}\}|}{\sum_{c'\in\mathcal{C}}|\{v \in c'(u)|(u,v) \in E \cup E_{\text{test}}\}|}$$

6 CONCLUSIONS

Prediction accuracy seems like a rather partial perspective for link prediction considering the new dimensions and role that social networks are acquiring, both as a service, a communication platform and a business. From the perspective of an effect on network evolution, prediction accuracy would just target network density by correctly predicting as many edges as possible. We find it natural to consider further qualities when setting the target for link prediction technology, taking into account that links are not necessarily all equally useful. In this paper we reflect on such perspectives and briefly lay out some possibilities in this direction.

We have implemented and tested the proposed metrics and, at the time of this writing, we are testing a wide array of link prediction algorithms [1,18,21,22,27,34] in order to empirically observe their behavior in light of the proposed dimensions, and analyze how the metrics relate (complement, correlate, etc.) to each other. We are also exploring how the discussed metrics can be explicitly targeted, either by devising new prediction algorithms that take the new dimensions into account, or by a post-prediction optimization of the output of an initial accuracy-oriented prediction algorithm, taking the desired metric as a second objective (by greedy re-ranking, multi-objective optimization, etc. [4,20]).

While the benefit of accuracy is easy to motivate, we still need to understand better what the network perspectives discussed here imply in terms of their desirability and value for the people in the network, or any other concerned party (platform

4

Track: 9th International Workshop on Modeling Social Media (MSM 2018)
Applying Machine Learning and AI for Modeling Social Media

WWW 2018, April 23-27, 2018, Lyon, France

owner, network data consumers, etc.). For instance, shortening distances seems good for everyone: anyone can reach more people through fewer introductions by common friends [26]. Connecting distant people exposes them to the risk of an enriching experience. Enhancing the degree equality or promoting long-tail users helps avoid the disengagement of less involved people, and the saturation of hubs. Weak links may alleviate social bubbles [28] and/or enhance the speed and diversity of the information flow through the network [10,37]. Exclusive links between communities may bring strategic value [12], and so forth. This is certainly domain dependent, but can probably be studied also at some level of abstraction, which we envisage as future work.

ACKNOWLEDGMENTS

REFERENCES

[1] L. Adamic and E. Adar. 2003. Friends and neighbors on the web. Social Networks 25, 3 (July 2003). Elsevier, Amsterdam, The Netherlands, 211–230.

[2] R. Agrawal, S. Gollapudi, A. Halverson and S. Ieong. 2009. Diversifying Search Results. In: *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining* (WSDM 2009). ACM, New York, NY, USA, 5–14.

[3] V. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre. 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics 10008, 10 (October 2008). IOP Publishing,Ltd, 1–12.

[4] P. Castells, N. Hurley and S. Vargas. 2015. Novelty and Diversity in Recommender Systems. In: *Recommender Systems Handbook, 2nd edition*, F. Ricci, L. Rokach, B. Shapira (Eds.). Springer, New York, NY, USA, Chapter 26, 881–919.

[5] O. Chapellle, J. Shihao, C. Liao, E. Velipasaoglu, L. Lai and S. Wu. 2011. Intent-based Diversification of Web Search Results: Metrics and Algorithms. *Information Retrieval* 14, 6 (December 2011). Springer, Vienna, Austria, 572–592.

[6] C. Clarke, M. Kolla, G. Cormack, O. Vechtomova, A. Ashkan, S. Buttcher and I. McKinnon. 2008. Novelty and diversity in Information Retrieval Evaluation. In: *Proceedings of the 31st ACM SIGIR International Conference on Research and development in Information Retrieval* (SIGIR 2008). ACM, New York, NY, USA, 659–666.

[7] A. Clauset, M. Newman and C. Moore. 2004. Finding Community Structure in Very Large Networks. *Physical Review E* 70, 6. APS, New York, NY, USA, 1–6.

[8] E. M. Daly, W. Geyer, and D. R. Millen. 2010. The Network Effects of Recommending Social Connections. In *Proceedings of the 4th Annual International ACM Conference on Recommender Systems* (RecSys 2010). ACM, New York, NY, USA, 301–304.

[9] P. Dankelmann, W. Goddard and C. Swart. 2004. The Average Eccentricity of a Graph and its Subgraphs. *Utilitas Mathematica* 65 (May 2004), Utilitas Mathematica Publishing Inc., Winnipeg, Manitoba, Canada.

[10] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. 2014. On Facebook, Most Ties are Weak. *Communications of the ACM* 57, 11 (November 2014). ACM, New York, NY, USA, 78–84.

[11] R. Dorfman. 1979. A Formula for the Gini Coefficient. The Review of Economics and Statistics 61, 1 (February 1979). The MIT Press, Cambridge, MA, USA, 146–149.

[12] M. Granovetter. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78, 6 (May 1973). University of Chicago Press, Chicago, IL, USA, 1360–1380.

[13] A. Goel, P. Gupta, J. Sirois, D. Wang, A. Sharma, and S. Gurumurthy. 2015. The Who-To-Follow System at Twitter: Strategy, Algorithms, and Revenue Impact. *Interfaces* 45, 1 (Febtuary 2015). INFORMS, Linthicum, MD, USA, 98–107.

[14] R. Guimerà, M. Sales-Pardo. 2009. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences* 106, 52 (December 2009). PNAS, Washington, DC, USA, 22073–22078.

[15] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh. 2013. WTF: The Who to Follow Service at Twitter. In *Proceedings of the 22nd Annual International Conference on World Wide Web* (WWW 2013). ACM, New York, NY, USA, 505–514

[16] I. Guy. 2015. Social Recommender Systems. In: *Recommender Systems Handbook, 2nd edition*, F. Ricci, L. Rokach, and B. Shapira (Eds.). Springer, New York, NY, USA, Chapter 15, 511–543.

[17] J. Hannon, M. Bennet, and B. Smyth. 2010. Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches. In *Proceedings of the 4th ACM Conference on Recommender Systems* (RecSys 2010). ACM, New York, NY, USA, 199–206.

[18] Y, Hu, Y. Koren, C. Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In: *Proceedings of the 8th IEEE International Conference on Data Mining* (ICDM 2008). IEEE, New York, NY, USA, 263–272.

[19] X. Huang, M. Tiwari, and S. Shah. 2013. Structural Diversity in Social Recommender Systems. In *Proceedings of the 5th ACM RecSys Workshop on Recommender Systems and the Social Web* (RSWeb 2013) *at the 7th ACM Conference on Recommender Systems* (RecSys 2013). CEUR Workshop Proceedings, Vol. 1066.

[20] N. Hurley and M. Zhang. 2011. Novelty and Diversity in Top-N Recommendation – Analysis and Evaluation. *ACM Transactions on Internet Technology* 10, 4 (March 2011). ACM, New York, NY, USA, 1-30.

[21] L. Katz. 1953. A new status index derived from sociometric analysis. Psychometrika 18, 1 (March 1953). Springer, New York, NY, USA, 39–43.

[22] E. Leicht, P. Holme and M. Newman. 2006. Vertex similarity in networks. *Physical Review E* 73, 2 (February 2006). APS, New York, NY, USA, 026120.

[23] D. Liben-Nowell and J. Kleinberg. 2007. The Link Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology* 58, 7 (May 2007). Wiley, Indianapolis, IN, USA, 1019-1031.

[24] R. Lichtenwalter, J. Lussier and N. Chawla. 2010. New perspectives and methods in link prediction. In: *Proceedings of the 16th ACM SIGKDD International conference on Knowledge Discovery and Data Mining* (KDD 2010). ACM, New York, NY, USA, 243–252.

[25] L. Lü and T. Zhou. Link Prediction in Complex Networks: A Survey. *Physica A* 390, 6 (March 2011). Elsevier, Amsterdam, The Netherlands, 1150–1170.

[26] M. Newman. 2010. *Networks. An Introduction.* Oxford University Press, Oxford, UK.

[27] X. Ning, C. Desrosiers and G. Karypis. 2015. A Comprehensive Survey of Neighborhood-Based Recommendation Methods. In: *Recommender Systems Handbook, 2nd edition*, F. Ricci, L. Rokach, and B. Shapira (Eds.). Springer, New York, NY, USA, Chapter 2, 37–76.

[28] E. Pariser. 2012. *The Filter Bubble.* Penguin Books, New York, USA.

[29] N. Parotsidis, E. Pitoura, and P. Tsaparas. 2016. Centrality-Aware Link Recommendations. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining* (WSDM 2016). ACM, New York, NY, USA, 503–512.

[30] M. Rosvall and C. T. Bergstrom. 2008. Maps of Random Walks on Complex Networks Reveal Community Structure. *Proceedings of the National Academy of Sciences*, 105, 4 (January 2008), National Academy of Sciences of the USA, Washington, DC, USA, 1118-1123

[31] J. Su, A. Sharma, and S. Goel. 2016. The Effect of Recommendations on Network Structure. *In Proceedings of the 25th Annual International Conference on World Wide Web* (WWW 2016), ACM, New York, NY, USA, 1157–1167.

[32] J. Tang, X. Hu and H. Liu. 2013. Social Recommendation: A Review. *Social Network Analysis and Mining* 3, 4 (December 2013). Springer, Vienna, Austria, 1113–1133.

[33] S. Vargas and P. Castells- 2011. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. In *Proceeedings of the 5th ACM International Conference on Recommender Systems* (RecSys 2011). ACM, New York, NY, USA, 109–116.

[34] S. White and P. Smyth. 2003. Algorithms for estimating relative importance in networks. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2003. ACM, New York, NY, USA, 266–275.

[35] J. Wu, G. Zhang and Y. Ren. 2017. A Balanced Modularity Maximization Link Prediction Model in Social Networks. *Information Processing & Management* 53, 1 (January 2017). Elsevier, Amsterdam, The Netherlands, 295–307.

[36] C. Zhai, W. Cohen and J. Lafferty. 2003. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In: *Proceedings of the 26st ACM SIGIR International Conference on Research and development in Information Retrieval* (SIGIR 2003). ACM, New York, NY, USA, 10–17.

[37] J. Zhao, J. Wu and K. Xu. 2010. Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E* 82, 1 (July 2010). APS, New York, NY, USA.

5