Enhancing Structural Diversity in Social Networks by Recommending Weak Ties

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

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

ABSTRACT

Contact recommendation has become a common functionality in online social platforms, and an established research topic in the social networks and recommender systems fields. Predicting and recommending links has been mainly addressed to date as an accuracy-targeting problem. In this paper we put forward a different perspective, considering that correctly predicted links may not be all equally valuable. Contact recommendation brings an opportunity to drive the structural evolution towards desirable properties of the network as a whole, beyond the sum of the isolated gains for the individual users to whom recommendations are delivered –global properties that we may want to assess and promote as explicit recommendation targets.

In this perspective, we research the definition of relevant diversity metrics drawing from social network analysis concepts, and linking to prior diversity notions in recommender systems. In particular, we elaborate on the notion of weak tie recommendation as a means to enhance the structural diversity of networks. In order to show the signification of the proposed metrics, we report experiments with Twitter data illustrating how state of the art contact recommendation methods compare in terms of our metrics; we examine the tradeoff with accuracy, and we show that diverse link recommendations result in a corresponding diversity enhancement in the flow of information through the network, with potential implications in mitigating filter bubbles.

CCS CONCEPTS

• **Information systems** → Recommender Systems; Online Social Networks; Evaluation of Retrieval Results • **Human-Centered Computing** → Social Network Analysis

KEYWORDS

contact recommendation; diversity; novelty; redundancy; social network analysis; weak tie; evaluation; metric

1 INTRODUCTION

Contact recommendation has gained importance as a key and pervasive functionality in large-scale online social networking industry [5,20,23,24,25]. It has likewise received increasing attention as a relevant and compelling scientific problem, also addressed in a related, link prediction perspective in the social network analysis field [32]. Predicting and recommending links has been mainly envisioned to date as an accuracy-targeting problem. From a recommendation perspective, prediction accuracy can be said to just target the network density by correctly predicting as many edges as possible. Density is obviously a major feature for a network to become a useful resource. Yet social networks have many further qualities that may enhance their value and performance for their users and businesses running on the network. Two predicted links being equally correct, one may enhance such properties more than the other.

It would seem natural to therefore consider what recommendation brings to the network from a wider perspective. Contact recommendation brings an opportunity to drive the structural evolution towards desirable properties of the network as a whole, beyond the sum of the isolated gains for the individual users to whom recommendations are delivered –global properties that we may want to assess and promote as explicit recommendation targets. Network science has produced a vast array of analysis and measurement tools to capture and understand the network structure and characteristics on manifold dimensions [36]. What network properties are desirable is certainly domain-dependent and specific to the purpose of the network.

In this paper we focus on the structural diversity of social networks as such a potential positive characteristic. We elaborate in particular on the notion of weak ties, largely studied in the social network analysis literature [4,9,10,21]. Starting from a generalized, structural definition of weak link [15], we consider both straightforward and more elaborate metric definitions. In order to test and observe how the proposed metrics behave, we run experiments with Twitter data where we compare the effect of state of the art contact recommendation algorithms in terms of the metrics. We further examine the accuracy tradeoff over greedy optimizations of the metrics at gradually aggressive diversification levels. Finally, we find grounding motivation for the proposed metrics by confirming their potential effect on the diversity in the information flow through the network, as a positive effect from the perspective of mitigating filter bubbles [37].

2 RELATED WORK

The problem of recommending contacts in social networks gained interest in the early 2000's as an application example of link prediction [1,33]. Many specific algorithms have been proposed for recommending people, based on the network structure around target users [24], user-generated content [25], or random walks to farther regions of the network [5,20,23]. These functionalities are

RecSys'18, October 2–7, 2018, Vancouver, BC, Canada.

ACM 978-1-4503-5901-6/18/10…$15.00
https://doi.org/10.1145/3240323.3240371

now present in mainstream social platforms, such as Twitter [23], Facebook or LinkedIn.

Link prediction was originally envisaged as a domain-independent problem relevant to many different types of complex networks besides the social domain [1,33,34]. From the recommender systems perspective, link prediction can be seen as a common recommendation task, where the predicted edges are delivered to users at either end of the links [5,23,24,25]. Taking the social network adjacency matrix as the rating matrix, any state of the art recommendation algorithm [25,27] can be applied straightforwardly, even unaware that users and items are the same thing. And many such methods indeed perform quite accurately [45]. Common recommender system evaluation methodologies and metrics [22,26] can be used as well, to assess the accuracy of predictions.

Research and development in link prediction and contact recommendation has been largely driven towards achieving accuracy, that is, maximizing the amount of recommended links that were present (though unobserved) in the network, or are useful and accepted by the target users. A few works have nonetheless considered the effects that the algorithms might have on the network on a wider perspective, thus providing a background for our work. Most authors in this scope have undertaken field studies of and working recommenders and algorithms from an external, "factual" stance.

Daly et al. [14] were among the first to bring attention to the action of recommendation on the global and local network structure. They observed how different recommendation algorithms had a diverse incidence on such network characteristics as the degree distribution skewness, node betweenness and triadic closure. In a similar spirit, Su et al. [41] analyzed the Twitter "Who To Follow" service, finding a popularity reinforcement effect in the degree distribution, along with an increase in triadic closure. Very recently, Aiello and Barbieri [3] studied the topological impact of the contact recommendation functionalities of Flickr and Tumblr at the local ego-network level, in terms of the popularity of linked nodes, and the overlap between recommendations as a primary measure of diversity. Huang et al [29] speculated on the interest of procuring diversity in recommendation; in a study with the LinkedIn "People You May Know" recommender, they found a tradeoff with recommendation accuracy, but a positive correlation with user engagement, using simple topological diversity metrics such as the number of recommended connected components and triangles. From a different take on demographic rather than structural diversity, Stoica et al [40] studied the potential inequality reinforcement effect of recommendation, and suggested a principled means to counter it.

Our research aims to take a step further from the observational analysis, by formalizing structural diversity metrics that can be systematically applied to evaluate and enhance recommendations from this perspective. We find a precedent of this aim in work by Parotsidis et al. [38], who also considered optimizing for network properties beyond accuracy, though their focus was on shortening distances –a different dimension from the one we are interested in. Moreover, beyond this, we seek connections between the proposed metrics and measurable positive effects on network functionalities, specifically in the transmission of information.

Diversity is also being the object of a sizeable body of research in the recommender systems field, leading to a wide an well-known variety of diversity notions, theories, metrics, and algorithms [2,12,28,44]. Our present research brings those perspective-widening views to the specific context where recommendation operates on social network structures, where diversity takes on new meanings and connections to network analysis concepts.

3 PRELIMINARIES

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 $\mathcal{U}$ is the set of people in the network, and $E \subset \mathcal{U}_*^2$ represents the set of relations between users in the network –$\mathcal{U}_*^2$ denoting the set of distinct user pairs. For each person $u \in \mathcal{U}$ we refer as $\Gamma(u)$ to her set of connections in the graph. Since we will consider directed networks, we shall differentiate between the incoming and outgoing neighborhoods $\Gamma_{in}(u)$ and $\Gamma_{out}(u)$.

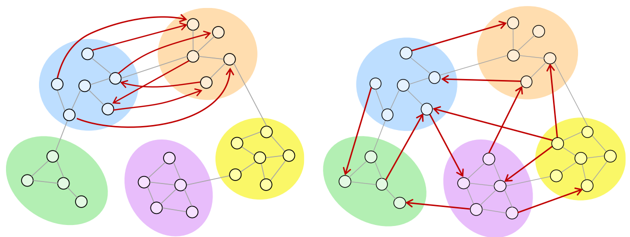
We state the contact recommendation problem for a given target user u as identifying a subset of users $\hat{\Gamma}_{out}(u) \subset \mathcal{U} \setminus \Gamma_{out}(u)$ that the user is likely to be interested in establishing a connection with. The accuracy of recommended contacts can be evaluated by common ranking-oriented metrics such as precision, recall, nDCG, based on the set of predicted links that actually satisfy the users, to which we can refer as the relevant recommendations. Relevance information can be obtained in different ways. In offline evaluation, a set of links can be hidden from the recommendation algorithms, and used as relevance ground truth, while the rest of links are fed to the algorithms as training data.

In order to evaluate recommended contacts beyond accuracy in terms of their effect on a network $\mathcal{G} = \langle \mathcal{U}, E \rangle$, we shall consider the extended network $\mathcal{G}' = \langle \mathcal{U}, E' \rangle$ that would result if users accepted all the recommended links, and added to them to the network. That is, $E' \leftarrow E \cup \hat{E}$ where $\hat{E} = \{(u,v) \in \mathcal{U}_*^2 | u \in \mathcal{U} \wedge v \in \hat{\Gamma}_{out}(u)\}$ is the set of all recommended edges. When considering the extended network, $\hat{\Gamma}_{out}(u)$ can be, typically, the top k users in the ranking of recommended contacts for u. Assuming all top k recommendations are going to be accepted is naturally quite strong and not realistic as a literal assumption; however, it helps envision and assess how the network may potentially change by the effect of a recommendation algorithm.

4 STRUCTURAL DIVERSITY

Diversity has been a major broad notion of interest in social network analysis, where it takes on different –often interrelated– meanings [9,15,21,42]. In this context, diversity has been related to such notions as the variety of people types in the network (e.g. cultural diversity, professional diversity, gender, age, etc.) [10], or the variety of relationships (family, friendship, work, etc.), involving side-information [7]. Further notions have been defined just in terms of the network topology, broadly related to the non-redundancy of network structures [9,15,21]. Social diversity has been found to bring benefits at different levels both for individuals and the network as a whole, ranging from information and innovation spread efficiency [15,42], to better professional and business performance [9,10], or even better health [7].

Most structural diversity notions revolve around the notion of weak tie [4,15,21]. Tie strength can be defined in terms of what



(a) Weak link redundancy (b) Weak link diversity

Figure 1. Redundant vs. spread weak tie distribution over communities.

the tie stands for in the domain where it is established, and how intense are the bond and interaction it represents: "the amount of time [involved in the relationship], the emotional intensity, intimacy (mutual confiding) and the reciprocal services which characterize the tie" [21]. From a domain-independent perspective, it is common to define the weakness of a link in terms of its structural properties [15,21], by assessing to what extent the link is not redundant. Inasmuch as weak links may foster novelty and efficiency, they can be a valuable complement of strong links, both for the individual and the network as a whole.

Link redundancy can be assessed in many different ways. For instance Granovetter related weakness to such sufficient conditions as: removing the link would break a connected component apart (global bridge), or the link connects people without common acquaintances (local bridge) [21]. These original notions were enticing and precursory; following Granovetter's lead, we find more informative generalizations for our purposes. In particular, De Meo et al [15] defined weak links as the ones that run across modular communities –a broader concept we will elaborate upon.

4.1 Global Diversity: Weak Ties

The weak tie notion we consider thus relies on a subdivision of the social network into communities. When this information is not explicit, a community detection algorithm can be used [13,18]. This makes any metric defined thereupon dependent on the detection method. De Meo et al. [15] showed that this weak tie definition is nonetheless fairly robust to the choice of detection algorithm.

A classical metric for quantifying the presence of weak ties given a community partition $\mathcal{C}$ of the network is the so-called modularity [13], which compares the number of links inside communities (strong ties) to the expected number we would find in a random network generated by a configuration model [36], that is, a random network keeping the same node degrees. Generalized to directed networks, it is defined as:

$$\operatorname{mod}(\mathcal{G}'|\mathcal{C}) = \frac{\sum_{u,v\in\mathcal{U}}(A_{uv} - |\Gamma_{out}(u)||\Gamma_{in}(v)|/|E'|)1_{c(u)=c(v)}}{|E'| - \sum_{u,v\in\mathcal{U}}(|\Gamma_{out}(u)||\Gamma_{in}(v)|/|E'|)1_{c(u)=c(v)}}$$

where $c(u)$ is the community user u belongs to, A is the adjacency matrix for the graph (that is, $A_{uv} = 1 \Leftrightarrow (u,v) \in E'$), and 1_{cond} is equal to 1 when $cond$ is true. Since low modularity indicates high diversity, we define the **modularity complement** (MC) metric as a linear transformation ranging in [0,1], where high and low values reflect high or low structural diversity, respectively:

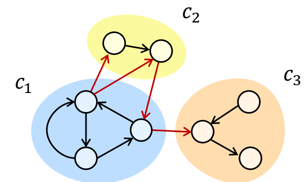


Figure 2. Community Edge Gini Complement example. In this case we have $n_{12} = 2$, $n_{13} = n_{21} = 1$, $n_{23} = n_{31} = n_{32} = 0$, $n_0 = 7$, therefore $X \to \langle 0,0,1,1,2,7\rangle$, which results in $\mathrm{CEGC}(\mathcal{G}'|\mathcal{C}) = 0.3739$.

$$\mathrm{MC}(\mathcal{G}'|\mathcal{C}) = \frac{1 - \operatorname{mod}(\mathcal{G}'|\mathcal{C})}{2}$$

A limitation of the MC metric is that it simply values the raw number of links crossing communities. It might be yet more interesting though to distinguish whether the weak links, valuable as they may be, are concentrated on a few pairs of communities, or are instead more widely distributed across many community pairs. Fig. 1 illustrates this issue. Based on this consideration, we refine the notion by De Meo et al. [15], by taking into account not just the redundancy involved in intra-community edges (as the weak tie notion does), but also the redundancy *between* weak ties, meaning whether they bridge the same or different communities. We hypothesize that the presence of many and well-distributed weak ties across many communities might be yet more beneficial at the global level.

A well-known measure of distribution fairness is the Gini index [17], which we can apply to the frequency (the distribution) of community pairs among the end points of weak ties, as a measure of how well-spread they are. Again, to have the metric orientation positively aligned with structural diversity, we shall take the complement of the Gini index, defined as follows. Given a community partition $\mathcal{C} = \{c_1, \ldots, c_m\}$, we count the set of edges crossing each two communities $c_i, c_j \in \mathcal{C}$:

$$n_{ij} = \left|\left\{(u,v) \in E' \middle| c(u) = c_i \wedge c(v) = c_j\right\}\right|$$

We count and sum, on the other hand, the number of all (structurally) strong ties –that is, the edges between people in the same community– $n_0 = \sum_{i=1}^{|\mathcal{C}|} n_{ii}$. Now we put together all these counts into $X = \left\{n_{ij} \middle| i \neq j\right\} \cup \{n_0\}$, and we sort X in increasing order $X \to \langle x_1, x_2, \ldots, x_N\rangle$. We then define the **community edge Gini complement** (CEGC) as the complement of the Gini index of X:

$$\mathrm{CEGC}(\mathcal{G}'|\mathcal{C}) = 1 - \frac{1}{N-1}\sum_{i=1}^{N}(2i - N - 1)\frac{x_i}{|E'|}$$

where $N = |X| = |\mathcal{C}|^2 - |\mathcal{C}| + 1$. Fig. 2 illustrates how the metric is computed. It is easy to understand that CEGC rewards a high number of weak links and low redundancy between them at the same time. The metric reaches value 0.0 when either all the links in the network run across the same two communities or there are no weak links at all; and value 1.0 when the same amount of edges crosses every pair of communities, and a few links (as many as the ones crossing any pair of communities) stay inside communities.

4.2 Local Redundancy: Transitive Closure

Complementary to the global diversity measures based on weak ties, we may consider triad closure as the smallest unit of structural redundancy. We may hence use the clustering coefficient for this purpose [36], which counts the ratio of closed triads in the network. Once again, in order to define a metric positively aligned with structural diversity levels, we take the **clustering coefficient complement** (CCC):

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

5 RECOMMENDATION EXPERIMENTS

In order to get an initial grasp of how the proposed metrics discern the behavior of common contact recommendation algorithms, and draw some insights and better understanding of what the metrics capture, we run a first experiment where we test them on two Twitter data samples.

5.1 Data and Task Setup

We run our experiment over dynamic interaction networks of Tweeter users, where an edge is present in the network $(u,v) \in E$ if u has replied, mentioned or retweeted a tweet posted by v. We download network data in a snowball sampling approach [19] using the pubic Twitter REST API as follows. Starting from a seed user, we extract a sample of tweets posted by that user. Then, we extract the set of outgoing interaction links induced by the retrieved tweets, that is, the set of users that were retweeted, replied or mentioned in the downloaded tweets. Those users are added to a queue, from which the next node is selected and the exploration continues until a desired number of users is retrieved.

We test two user tweet sampling approaches: a) all tweets posted by each user in a fixed period of time, and b) the n most recent tweets of each user. We thus obtain two datasets: one based on all tweets posted by the retrieved users between 16$^{\text{th}}$ June and 16$^{\text{th}}$ July 2015, to which we shall henceforth refer as the "1 month dataset"; and one based on the last 200 tweets posted by the users, to which we shall refer as the "200 tweets dataset". Table 1 shows the size and details of the obtained datasets.

We evaluate the accuracy of recommendations in an offline approach based on a temporal network split into training data (which is supplied as input to the contact recommendation algorithms) and test data (which is used to assess which recommendations match links that were created in the test period). In the 1 month dataset, we take the interactions defined by tweets posted in the three first weeks (until 9$^{\text{th}}$ July 2015) as the training set, and the remaining links as the test set. In the 200 tweets dataset, interactions included in the first 80% of the tweets define the training graph, and the rest the test graph. If the same link appears in both training and test, it is removed from the test network. Table 1 shows the network split details. For relevance-oriented accuracy metrics such as precision or recall [6] a recommended contact v is considered relevant for a target user u if an edge (u,v) is present in the test network.

In our experimental setup, we exclude from recommendations all the reciprocating links, that is, we do not recommend people who follows back the target user. There are two reasons for this: first, social platforms like Twitter already notify users when someone starts following them, so the recommendation would be redundant; second, the reciprocity ratio in Twitter is very high, and hence would trivialize recommendations and distort the accuracy evaluation.

As explained in section 4.1, community-based metrics require a network subdivision into communities. We tested different community detection algorithms for that purpose and, consistently with observations by De Meo et al. [15], we found that the comparison between recommendation approaches was broadly insensitive to the choice of community detection algorithm. We therefore report results using just one of the tested methods, namely the Louvain algorithm [8], one of the most widely used and effective in the community detection literature.

5.2 Recommendation Algorithms

We test our metrics on 8 different contact recommendation algorithms, selected based on their popularity and performance [39], which we can classify in five groups:

- **Neighborhood-based:** We implement three common algorithms based on different measures of neighborhood overlap between the target and the recommended user: Most common neighbors (MCN) based on the plain neighborhood intersection [35], the Jaccard neighborhood similarity [30], and the Adamic-Adar coefficient [1,33].
- **Random walks:** We implement the personalized SALSA algorithm applied in the Twitter 'Who-to-follow' system [20,23,32], only without restricting the circle of trust, for simplicity.
- **Content-based algorithms:** As a representative option, we use the cosine similarity between users as the ranking function [25], where users are represented by the centroid of the *tf-idf* vectors [6] of the tweets posted by their neighbors. Whether the neighborhood includes outgoing, incoming links, or both, is a configuration option of the algorithm, which we tune in the same way as any other parameters.
- **Adaptation of recommender system algorithms:** We adapt a top-performing matrix factorization algorithm (Implicit MF) [27] to the task of recommending users.
- **Reference baselines:** We include two trivial non-personalized approaches, most-popular and random recommendation, as a sanity-check and rock-bottom reference.

We tune the free parameters for each algorithm by a simple grid search targeting P@10. In the MCN, Jaccard, Adamic-Adar, and content-based algorithms, we also selected the optimal neighborhood direction (in, out, undirected) for the target and candidate users.

5.3 Accuracy and Structural Diversity

Along with the parameter settings for the tested contact recommendation algorithms, Table 2 shows the accuracy and structural diversity values at cut-off 10. As explained in section 3, for structural metrics this cutoff means that the top 10 recommended by each algorithm are added to the training network $\mathcal{G}$ to form the extended network $\mathcal{G}'$, on which the metrics are computed.

Table 1. Dataset details.

	Complete network		Training network			Test network	
Dataset	**#Users**	**#Edges**	**#Users**	**#Edges**	**#Communities**	**#Users**	**#Edges**
1 month	10,019	234,869	9,528	170,425	8	7,902	57,846
200 tweets	10,000	164,653	9,985	137,850	10	5,652	21,598

Table 2. Accuracy and structural diversity of state of the art contact recommendation algorithms. The recommenders are ordered by decreasing P@10. The cells for precision and recall are colored from white (lowest values) to blue (higher values), and structural diversity metrics from red, when the value is lower than in the training graph, to blue when higher. The highest value of each column is highlighted with bold font. All the pairwise differences between algorithms in precision and recall are statistically significant (2-tailed Student t-test at $p < 0.05$), except Adamic-Adar and Implicit MF on the 200 tweets dataset. Structural metrics yield a single global value rather than an average per user, hence common statistical significance tests are not directly applicable.

	Recommender	Parameter settings	P@10	Recall@10	MC	CEGC	CCC
1 month	Implicit MF	$k = 260, \lambda = 150, \alpha = 40$	**0.0625**	**0.1060**	0.1550	0.0447	0.9766
	Personalized SALSA	Authorities, $\alpha = 0.99$	0.0577	0.0990	0.1656	0.0447	0.9819
	Adamic-Adar	und, in, und	0.0505	0.0697	0.1487	0.0413	0.9748
	MCN	und, in	0.0476	0.0647	0.1461	0.0403	0.9746
	Popularity	-	0.0234	0.0409	**0.2947**	0.0613	**0.9890**
	Jaccard	und, in	0.0169	0.0209	0.1464	0.0434	0.9652
	Centroid CB	in	0.0156	0.0198	0.1652	0.0498	0.9627
	Random	-	0.0006	0.0009	0.2797	**0.0901**	**0.9839**
	Training graph	-	-	-	*0.1464*	*0.0390*	*0.9829*
200 tweets	Implicit MF	$k = 300, \lambda = 150, \alpha = 40$	**0.0236**	**0.0589**	0.2132	0.1326	0.9520
	Adamic-Adar	und, in, und	0.0233	0.0540	0.2076	0.1180	0.9447
	MCN	und, in	0.0222	0.0499	0.2048	0.1138	0.9433
	Personalized SALSA	Authorities, $\alpha = 0.99$	0.0208	0.0516	0.2369	0.1412	0.9594
	Centroid CB	in	0.0157	0.0333	0.2154	0.1251	0.9182
	Jaccard	und, in	0.0132	0.0306	0.2041	0.1195	0.9065
	Popularity	-	0.0098	0.0221	**0.3371**	0.1559	**0.9822**
	Random	-	0.0003	0.0007	0.3317	**0.2276**	0.9795
	Training graph	-	-	-	*0.2081*	*0.1134*	*0.9559*

Consistently with prior experiments [39], we observe that Implicit MF stands out above the rest of the algorithms in all datasets in terms of accuracy, followed by Adamic-Adar and MCN. In terms of structural diversity metrics, the top approaches are not personalized: popularity and random recommendation. This is to be expected: creating weak links outside communities is much more likely when the topological placement of the target user is simply ignored by the non-personalized algorithm.

A closer look at these two recommendations is illustrative of the nuances that distinguish the diversity metrics. Random recommendation trivially optimizes CEGC: a fair spread of links across communities is quite likely when the links are placed uniformly at random. Popularity in contrast concentrates links on fewer communities: the ones the top most popular users belong to, hence incurring in some redundancy in comparison. However, random links seem to fall inside communities slightly more often (in raw numbers) than linking to the same few popular users, as reflected by the MC metric. Finally, CCC shows that adding links to highly connected people seems to close very few triangles –slightly less than random. Network hubs have many neighbors that are not shared by the average target user; those are triads that are not transitively closed.

Among personalized algorithms, SALSA produces the most diverse network structures in terms of the three metrics. Implicit MF also stands out, especially in the clustering coefficient complement. Adamic-Adar, MCN and Jaccard, unsurprisingly, obtain low values for MC and CCC, since these methods favor people with many common neighbors, thus likely closing triads, and likely within the same community. Jaccard is particularly notorious in CCC, as closing as many triads as possible is literally the objective of this recommender. Finally, the content-based approach illustrates to what extent the metrics may measure complementary nuances: while this algorithm seems to produce a fair number of well-distributed weak links according to MC and CEGC, it seems to close quite an amount of transitive triads as captured by CCC.

6 EFFECTS ON INFORMATION DIVERSITY

We have so far observed how the proposed metrics describe the effect of contact recommendation on the structural diversity of networks. In order to assess to what extent the proposed perspective may find deeper implications, we analyze how the structural dimension of contact recommendation may have consequences in the behavior of the network, in particular, in the diffusion of information, one of the foremost functionalities of online social networks. We aim to test, in particular, the hypothesis that the more structurally diverse is the recommendation, the more diverse and

ALGORITHM 1: Global Greedy Reranking

Input: $\hat{E} \subset \mathcal{U}_*^2$ original recommendations
$f: \hat{E} \to \mathbb{R}$ original recommendation ranking function
μ metric to optimize
k diversification cutoff
$\lambda \in [0,1]$ degree of diversification
$\mathcal{G} = \langle \mathcal{U}, E \rangle$ training graph
Output: $\mathcal{S}$ modified recommendations (a set of ordered lists)
begin
  $\mathcal{S} \leftarrow \mathrm{sort}(\hat{E}, f)$ // Edges are grouped by source node and sorted by f
  for $u \in \mathcal{U}$ **do**
    for $i \leftarrow 1$ **to** k **do**
      $j_0 \leftarrow \arg\max_{j:k<j\le|\mathcal{S}_u|} \phi(j|\mathcal{S},u,i,f,\mu,\lambda)$ // $\mathcal{S}_u \equiv$ ranking for user u in $\mathcal{S}$
      if $\phi(j_0|\mathcal{S},u,i,f,\mu,\lambda) > \phi(i|\mathcal{S},u,i,f,\mu,\lambda)$ **then** $\mathrm{swap}(\mathcal{S}_u, i, j_0)$
  return $\mathcal{S}$
end

Function $\phi(j|\mathcal{S},u,i,f,\mu,\lambda)$ // The dual objective function
begin
  return $\lambda\ \mathrm{norm}\big(f(\mathcal{S}_u[j])\big) + (1-\lambda)\ \mathrm{norm}\Big(\mu\big(\mathcal{G}'_{\mathcal{S}\langle u:i/j\rangle@k}\big)\Big)$
end

novel (non-redundant) will be the information flow through the network by consequence of recommendation.

6.1 Structural Diversity Enhancement

In order to check how diversity metrics correlate with information flow properties, we start by defining a procedure to optimize a given diversity metric by gradual aggressivity levels. As a common approach, starting with a well-performing baseline contact recommendation algorithm (in terms of accuracy), we apply a greedy reranking that partially targets the metric of interest [12,43]. While typical greedy optimization procedures target a local objective that depends on the ranking for a single user [11], in our case the targeted metrics depend on all recommendations together.

We therefore define the global optimization problem as follows. Given a recommendation $\hat{E}$ with rankings defined by a function $f: \hat{E} \to \mathbb{R}$ (the recommender "score"), we seek new rankings that balance the accuracy achieved by the initial recommendation, and the targeted diversity metric. Inasmuch as we consider a cutoff of interest k for the diversity metric, defining this new ranking amounts to selecting a subset $\mathcal{S} \subset \hat{E}$, where $\mathcal{S} = \bigcup_{u\in\mathcal{U}} \mathcal{S}_u$ consists of subsets $\mathcal{S}_u \subset \{u\} \times \hat{\Gamma}_{out}(u)$ of size $|\mathcal{S}_u| = k$ for each user. It is common to combine this dual goal by targeting a linear combination of the two objectives (original ranking –for accuracy– plus diversity), which we can express as:

$$\mathcal{S} = \underset{S\subset\hat{E}:|S_u|=k}{\arg\max}\ (1-\lambda)\sum_{u\in\mathcal{U}}\ \sum_{(u,v)\in S_u} f(u,v) + \lambda\,\mu(\mathcal{G}'_S)$$

where μ is the targeted diversity metric, and $\mathcal{G}'_S = \langle \mathcal{U}, E \cup S\rangle$ denotes the extension of a network $\mathcal{G} = \langle \mathcal{U}, E\rangle$ by adding the links in a recommendation subset S. The parameter λ adjusts how aggressive is the reranking, from no change at $\lambda = 0$, to ignoring the initial ranking at $\lambda = 1$.

Having set the objective, our greedy procedure works as follows (see Algorithm 1). For each individual ranking S_u, starting from the top down, we consider swapping the i-th element in the top k of S_u with the user at position j below k that maximizes the following greedy objective function:

$$\phi(j|S,u,i,f,\mu,\lambda) = \lambda\ \mathrm{norm}\big(f(S_u[j])\big) + (1-\lambda)\ \mathrm{norm}\Big(\mu\big(\mathcal{G}'_{S\langle u:i/j\rangle@k}\big)\Big)$$

where $S\langle u: i/j\rangle$ denotes swapping the i-th and j-th elements in S_u, and $@k$ indicates taking the subset of elements ranked in the top k of each ranking $S_v \subset S$ for all users $v \in \mathcal{U}$. This best swap is applied only if $\phi(j|S,u,i,f,\mu,\lambda) > \phi(i|S,u,i,f,\mu,\lambda)$, that is, if the target function is improved. As is common, we normalize the two components of the linear combination in ϕ, as expressed by $\mathrm{norm}(\cdot)$, to make their aggregation meaningful [43]. We use the rank-sim normalization scheme [31] for this purpose.

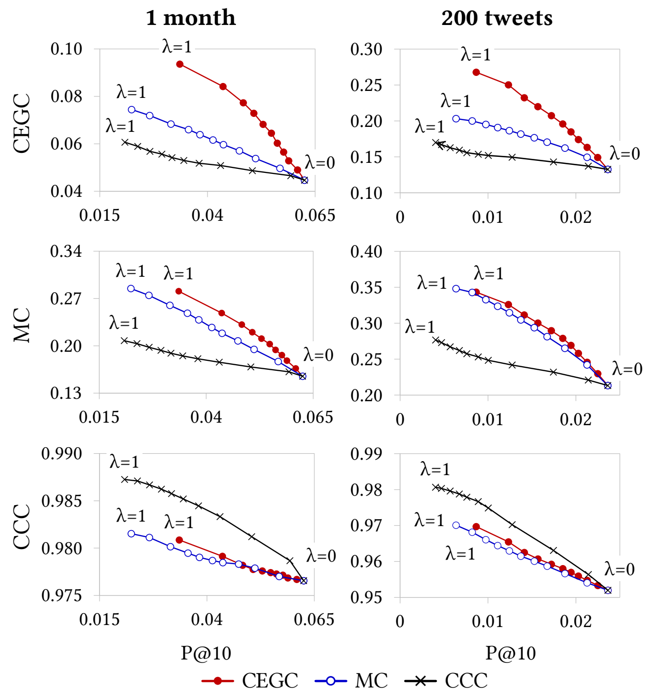


Figure 3. Plot of diversity (y axis) against P@10 (x axis) for the three diversity metrics, and the corresponding rerankers. Points on the curves represent reranked recommendations for λ ranging from 0 (undiversified initial recommendation) to 1 (maximum diversity), by increments of 0.1.

Fig. 3 shows the effect of reranking for the three diversity metrics defined in section 4. As initial recommendations we take the top 100 contacts returned by Implicit MF (the most accurate in our experiments in section 5), and we target the metrics at a $k = 10$ cutoff. For the optimization of MC, we simplify the greedy diversity objective to just count whether the recommended link is weak or not. For CEGC, we require the edge $S[j]$ to be weak as an additional condition for a swap to be applied in the reranking procedure.

The figure shows the clear and typical diversity-accuracy tradeoff [12] in all cases. We also see that rerankings are consistent in that each is best at enhancing the metric it optimizes for –not just on the diversity axis, but also in the tradeoff with accuracy. We also see that the MC and CEGC optimizers have a comparable effect in both CCC and even MC, which makes sense as both diversifiers promote a similar number of weak links, as one should expect. However, the graphics for the CEGC metrics (bottom of figure) evidence our point as to what this metric aims to capture: not all weak ties are equally weak, which plain modularity optimization fails to take into proper account.

6.2 Information Diffusion: Novelty, Diversity and Speed

We analyze the effect of structurally diverse recommendations on information spread based on the messages (tweets) that users post and pass on in the network. We shall observe three main dimensions in the message flow: speed, novelty, and diversity.

Though not the primary focus of our analysis, we keep an eye on the speed of propagation, as one of the most commonly analyzed network efficiency features in diffusion processes [16]. We measure it as the sum of the number of tweets received by each user in the network at a given moment in time:

$$\text{speed}(t) = \sum_{u \in \mathcal{U}} |\mathcal{M}_u(t)|$$

where $\mathcal{M}_u(t)$ denotes the set of messages (tweets) that user u has received since the diffusion started to be measured, until time t.

In order to measure information novelty and diversity we use tweet hashtags as a readily available indicator of information topicality. For this purpose, we shall represent tweets as sets of tags: the ones the tweet contains. If $\mathcal{H}$ denotes the set of all hashtags, then tweets are subsets $i \subset \mathcal{H}$. Based on this representation, we assess the novelty a tweet carries for a user who receives it simply by the number of tweet hashtags that the user had never used herself before. Specifically, we define the **external hashtag rate** (EHR) of the information flow as:

$$\text{EHR}(t) = \frac{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{M}_u(t)} |i \setminus \mathcal{H}_u^0(t)|}{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{M}_u(t)} |i|}$$

where $\mathcal{H}_u^0(t) = \bigcup_{i \in \mathcal{M}_u^0(t)} i$ denotes the set of tags used by u in the tweets she posted up to time t –denoted as $\mathcal{M}_u^0(t)$.

As to diversity, we consider that the information is diverse to the extent that different hashtags are evenly distributed over the population through the tweets. We resort again to the (complement of the) Gini index as a measure of [17], and define the **hashtag Gini complement** (HGC) as:

$$\text{HGC}(T) = 1 - \frac{1}{|\mathcal{H}| - 1} \sum_{j=1}^{|\mathcal{H}|} (2j - |\mathcal{H}| - 1)\, p\big(h_j|T\big)$$

where h_j denotes the j-th least frequently spread hashtag, and:

$$p(h|T) = \frac{|\{u \in \mathcal{U} | h \in \mathcal{H}_u(t)\}|}{\sum_{h^* \in \mathcal{H}} |\{u \in \mathcal{U} | h^* \in \mathcal{H}_u(t)\}|}$$

where $\mathcal{H}_u(t) = \bigcup_{i \in \mathcal{M}_u(t)} i$ denotes the set of tags used by u in the tweets she received.

6.3 Diffusion Procedure

In order to analyze how the recommendation can impact information diffusion, we simulate the publication and transmission of information through tweets and retweets over a network $\mathcal{G}'$ extended by different recommendations. We base the simulation on an actual set of tweets –the same we have used in the experiments reported in the previous sections– where for every user we have a list of tweets she has posted, and a list of tweets she has retweeted.

Table 3. Data description for the information diffusion experiment.

Dataset	#Tweets	#Hashtags (unique)
1 month	87,837	110,578 (1115)
200 tweets	21,513	24,623 (378)

We follow an iterative simulation procedure that roughly mimics the main information passing actions on Twitter. At each step (a "time point") every user randomly selects (without replacement) and posts one of her authored tweets, which is passed on to all the user's incoming neighbors, who add it to a personal inbox (a timeline) of received tweets. Next, every user retweets all the received tweets they find interesting, which are in turn received by all users linking to her. We consider a user would find a tweet worth retweeting if she actually retweeted it for real sometime according to our downloaded Twitter data. Hence at every time point, each user posts one original tweet, and retweets a few tweets, as long as any are available on her lists. When a tweet is received more than once by the same user, it is simply ignored (i.e. only the first time counts, as on Twitter). The simulation finally stops when all original tweets have been posted and all inboxes empty up.

Note that the way our simulation is configured makes it actually deterministic for our purposes. Since it applies an exhaustive propagation, it is equivalent to a set of graph explorations, one for each tweet in the dataset, starting at the user who authored it, following all incoming links backwards, and stopping every time a user is reached who had not retweeted the tweet in our dataset. The resulting search tree for each tweet covers the set of users who receive the tweet, which is all we need to measure the outcomes of interest for our study. The simulation perspective enables a smooth generalization of our analysis to many other communication protocols one may wish to explore.

Once we determine the set of tweets that have reached every user, we can assess information properties (volume, diversity, novelty) by such metrics as defined in the previous section, based on the tweet hashtags.

6.4 Results

We finally check the effect of structural diversity on the information flow by running the three rerankers, targeting MC, CEGC and CCC respectively, over the Implicit MF baseline with different intensity of diversification, adjusted by the λ parameter. For each reranked recommendation (3 rerankers × 11 values of λ, adding to 33 recommendations for each dataset), we run the publication and propagation of tweets over the network extended by the corresponding recommendation, as explained in the previous section, and we measure the resulting speed, novelty and diversity of the information flow.

As previously mentioned, we use in this experiment the same tweets we downloaded to build a network sample as described in section 5.1. In particular, we use the test tweets (the ones posted after the temporal split point) containing at least one hashtag

–they otherwise have no impact on our diversity measures. Moreover, to avoid noise and a heavy-tail distortion, we consider only hashtags appearing in at least 25 tweets. Details on the number of tweets and hashtags used in our experiment are shown in Table 3.

Fig. 4 shows the results, as a tradeoff between information flow enhancement and the accuracy of the recommended contacts. We first of all confirm that the metric that most enhances information diversity is CEGC. To the effectiveness of weak links promoted by MC, CEGC adds a further redundancy reduction by more carefully selecting the weak ties to be promoted. In the 1 month dataset, both strategies procure a similar level of information diversity at the most aggressive structural diversification level, but at the expense of an additional accuracy loss in the case of MC. In the 200 tweets dataset MC underperforms also in terms of diversity. CCC, on its side, seems to have a more erratic effect overall, though it seems to achieve some diversity enhancement at the most aggressive levels, by an additional accuracy penalization. It therefore seems advantageous to take community structure into account, which CCC does not.

In terms of novelty there seems to be no clear winner: the three strategies enhance this property to a quite similar degree with slight differences. This is consistent with Granovetter's theory, stating that weak links are a source of novel information for people [21]. On the other hand, the structural metrics do not seem to display a clear effect in terms of propagation speed. A slowdown is observed on the 200 tweets data, and mixed effects are seen on the 1 month dataset. This shows that diversity and novelty are not a trivial reflection of the information volume.

Overall, we may say that enhancing the community edge Gini complement provides the best trade-off in terms of diffusion properties, structural properties and accuracy.

7 CONCLUSIONS

Considering the broad effect that recommending new edges can have on the evolution of a social network, beyond just increasing their density, seems only natural –more so when massive online social networks may me approaching some stage of maturity, be it just in terms of their sheer size, towards a stationary, slower growth. Inasmuch as contact recommendation may play a substantial part in shaping the network growth, an opportunity lies therein to steer the network evolution towards enhancing profitable properties. At a general level, structural diversity is one of the widely analyzed characteristics of social networks that have been the object of broad study and some speculation as to its implications on the performance of networks and their members, and the benefits for the common good from different perspectives [4,7,9,10,21,42]. We explore a particular angle to this outlook here, by focusing on diversity notions related to specific definitions of edge strength and redundancy.

We show that it is possible to define sound recommendation evaluation metrics based on these concepts, and optimization strategies that consistently enhance structural diversity by gradually retargeting recommendations towards the corresponding metrics, allowing to keep the tradeoff with accuracy at the desired level. We further find a practical signification for the proposed recommendation metrics in the positive effect the characteristics they

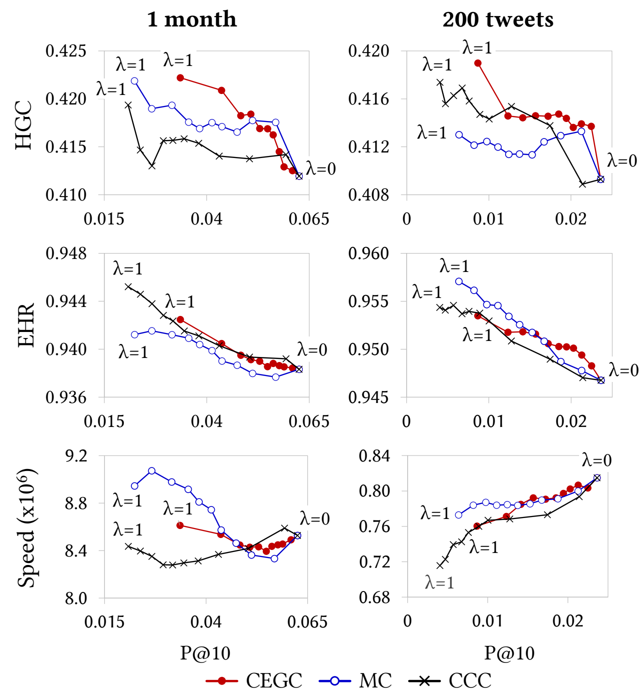


Figure 4. Information flow diversity, novelty and efficiency (y axis) against recommendation accuracy (P@10, x axis) for the diversity-targeting rerankers. Arrows show the direction of the curves from $\lambda = 0$ to $\lambda = 1$.

measure can have on the behavior of networks as channels for the dissemination of relevant information. We focus in particular on information diversity and novelty as positive factors in healthy social media and filter bubble mitigation [37]. We introduce a novel distinction in the concept of weak links that translates into higher specific enhancements in information diversity, compared to previously considered structural definitions of weak tie [15,21].

The directions to continue our research are manifold. Many further metrics and notions from social network analysis can be adapted to evaluate contact recommendation from the corresponding perspective. Any global objective property found to be desirable for networks could be amenable to a similar development as we have undertaken here. Identifying such properties is in fact a worthy research goal in itself. In the impact on information spread, further communication dynamics can be considered besides the Twitter model, and different information features, diversity variants and spaces (e.g. opinion diversity) can be studied. The effect of structural recommendation properties on further network functionalities and dynamic processes beyond information diffusion can be envisaged as well, such as network growth, robustness, or influence propagation. Moving beyond the pure accuracy objective in contact recommendation has barely started to be explored, and seems like a sensible direction to pursue.

ACKNOWLEDGMENTS

This work was partially supported by the national Spanish Government (grant nr. TIN2016-80630-P).

REFERENCES

[1] Lada A. Adamic and Eytan Adar. 2003. Friends and Neighbors on the Web. *Social Networks*, 25, 3 (July 2003), 211-230.

[2] Panagiotis Adamopoulos and Alexander Tuzhilin. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology*. 5, 4 (December 2014).

[3] Luca M. Aiello and Nicola Barbieri. 2017. Evolution of Ego-networks in Social Media with Link Recommendations. In *Proceedings of the 10th ACM International Conference on Web Search and Data Mining* (WSDM 2017). ACM, New York, NY, USA, 111-120.

[4] Sinan Aral. The Future of Weak Ties. 2016. *American Journal of Sociology*, 121, 6 (May 2016), 1931-1939.

[5] Lars Backstrom and Jure Leskovec. 2011. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the 4th ACM international conference on Web search and data mining* (WSDM 2011). ACM, New York, NY, USA, 635-644.

[6] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 2011. *Modern Information Retrieval: The Concepts and Technology Behind Search* (2nd ed.). Addison-Wesley Publishing Company, USA.

[7] John. C. Barefoot, Morten Grønbæk, Gorm Jensen, Peter Schnohr, and Eva Prescott. 2005. Social Network Diversity and Risks of Ischemic Heart Disease and Total Mortality: Findings from the Copenhagen City Heart Study. *American Journal of Epidemiology* 161, 10 (May 2005), 960–967.

[8] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambriotte and Etienne Lefebvre. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* 10 (October 2008).

[9] Ronald. S. Burt. 1995. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA, USA.

[10] Ronald. S. Burt. 2004. Structural holes and good ideas. *American Journal of Sociology* 110, 2 (September 2004), 349–399.

[11] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 1998). ACM, New York, NY, USA, 335-336.

[12] Pablo Castells, Neil J. Hurley and Saúl Vargas. 2015. Novelty and Diversity in Recommender Systems. In F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook* (2nd ed.). Springer, New York, NY, USA, 881-918.

[13] Aaron Clauset, Mark E.J. Newman and Christopher Moore. 2004. Finding Community Structure in Very Large Networks. *Physical Review E*, 70, 6 (December 2004).

[14] Elizabeth M. Daly, Werner Geyer, and David R. Millen. 2010. The network effects of recommending social connections. In *Proceedings of the 4th ACM Conference on Recommender Systems* (RecSys 2010). ACM, New York, NY, USA, 301-304.

[15] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2014. On Facebook, most ties are weak. *Communications of the ACM* 57, 11 (October 2014), 78-84

[16] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. 2012. Why rumors spread so quickly in social networks. *Communications of the ACM* 55, 6 (June 2012), 70-75.

[17] Robert Dorfman. 1979. A Formula for the Gini Coefficient. *The Review of Economics and Statistics*, 61, 1 (February 1979), 146-149.

[18] Santo Fortunato. 2010. Community Detection in Graphs. *Physics Reports*, 486, 3-5 (February 2010), 75-174.

[19] Leo A. Goodman. 1961. Snowball Sampling. *Annals of Mathematical Statistics* 31, 1 (March 1961), 148-170.

[20] Ashish Goel, Pankaj Gupta, John Sirois, Dong Wang, Ameesh. Sharma, and Siva Gurumurthy. 2015. The Who-To-Follow System at Twitter: Strategy, Algorithms, and Revenue Impact. *Interfaces* 45, 1 (February 2015), 98–107.

[21] Mark S. Granovetter. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78, 6 (May 1973), 1360-1380.

[22] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook* (2nd ed.). Springer, New York, NY, USA, 265-309.

[23] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. WTF: the Who to Follow Service at Twitter. In *Proceedings of the 22nd international conference on World Wide Web* (WWW 2013). ACM, New York, NY, USA, 505-514.

[24] Ido Guy. 2015. Social Recommender Systems. In F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook* (2nd ed.). Springer, New York, NY, USA, 511-543.

[25] John Hannon, Mike Bennett, and Barry Smyth. 2010. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the 4th ACM Conference on Recommender Systems* (RecSys '10). ACM, New York, NY, USA, 199-206.

[26] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22, 1 (January 2004), 5-53.

[27] Yifan Hu, Yehuda Koren and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In: *Proceedings of the 8th IEEE International Conference on Data Mining* (ICDM 2008). IEEE Computer Society, Washington, DC, USA, 15-19.

[28] Neil Hurley and Mi Zhang. 2011. Novelty and Diversity in Top-N Recommendation -- Analysis and Evaluation. *ACM Transactions on. Internet Technology* 10, 4 (March 2011).

[29] Xinyi L. Huang, Mitul Tiwari, and Sam 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).

[30] Paul Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 142 (January 1901), 547-579.

[31] Joon Ho Lee. 1997. Analyses of multiple evidence combination. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 1997). ACM, New York, NY, USA, 267-276.

[32] Ronny Lempel and Shlomo Moran. 2001. SALSA: the stochastic approach for link-structure analysis. *ACM Transactions on Information Systems* 19, 2 (April 2001), 131-160.

[33] David Liben-Nowell and Jon Kleinberg. 2003. The link prediction problem for social networks. In *Proceedings of the 12th International Conference on Information and Knowledge Management* (CIKM '03). ACM, New York, NY, USA, 556-559.

[34] Liyuan Lü and Tao Zhou. 2010. Link Prediction in Complex Networks: A Survey. *Physica A* 390, 6 (March 2011), 1150-1170

[35] Mark E.J. Newman. 2001. Clustering and preferential attachment in growing networks. Physical Review E 64 (July 2001).

[36] Mark E. J. Newman. 2010. *Networks: An Introduction*. Oxford University Press, Oxford, UK.

[37] E. Pariser. 2011. *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*. Penguin Press, 2011.

[38] Nikos Parotsidis, Evaggelia Pitoura, and Panayiotis 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.

[39] Javier Sanz-Cruzado and Pablo Castells. 2018. Contact Recommendations in Social Networks. In: I. Cantador, S. Berkovsky, D. Tikk (Eds.), *Collaborative Recommendations: Algorithms, Practical Challenges and Applications*. World Scientific Publishing, Singapore.

[40] Ana-Andreea Stoica, Christopher Riederer and Augustin Chaintreau. 2018. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In *Proceedings of the 2018 World Wide Web Conference* (WWW 2018). IW3C, Geneva, Switzerland, 923-932.

[41] Jessica Su, Aneesh Sharma, and Sharad Goel. 2016. The Effect of Recommendations on Network Structure. In *Proceedings of the 25th International Conference on World Wide Web* (WWW 2016). IW3C, Geneva, Switzerland, 1157-1167.

[42] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109, 16 (April 2012), 5962–5966.

[43] David Vallet and Pablo Castells. 2012. Personalized diversification of search results. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 2012). ACM, New York, NY, USA, 841-850

[44] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2nd ACM Conference on Recommender Systems* (RecSys 2008). ACM, New York, NY, USA, 123-130.