Voting in Social Networks

Paolo Boldi[1]* Francesco Bonchi[2] Carlos Castillo[2] Sebastiano Vigna[1]*

[1]DSI, Università degli Studi di Milano, Italy
{boldi,vigna}@dsi.unimi.it

[2]Yahoo! Research Labs Barcelona, Spain
{bonchi,chato}@yahoo-inc.com

ABSTRACT

A voting system is a set of rules that a community adopts to take collective decisions. In this paper we study voting systems for a particular kind of community: electronically mediated social networks. In particular, we focus on *delegative democracy* (a.k.a. proxy voting) that has recently received increased interest for its ability to combine the benefits of direct and representative systems, and that seems also perfectly suited for electronically mediated social networks. In such a context, we consider a voting system in which users can only express their preference for one among the people they are explicitly connected with, and this preference can be propagated transitively, using an attenuation factor.

We present this system and we study its properties. We also take into consideration the problem of missing votes, which is particularly relevant in online networks, as some recent case shows. Our experiments on real-world networks provide interesting insight into the significance and stability of the results obtained with the suggested voting system.

Categories and Subject Descriptors H.4.3 [**Information Systems Applications**]: Communications Applications
General Terms Algorithms
Keywords Social networks, e-democracy

1. INTRODUCTION

On April 23, 2009, Facebook announced the preliminary results of a site-governance vote, a ballot in which the users were called to express themselves on a change in the terms of use of the popular social networking site. The event was presented as an important step in the history of social networking[1].

The vote came as the result of a petition by thousands of outraged Facebook users, who criticized the social networking site for claiming too many rights over the user-generated content. In an attempt to make the social network more democratic, Facebook decided to let the users choose between the current terms of use and two new documents it had published regarding site governance. It was also announced that at least 30% of the roughly 200 millions active Facebook users would have had to vote for the results to be binding. The outcome was that the new rules were preferred by 74.4% of the voters. And while only 600,000 users (0.3%) voted—a turnout far smaller than Facebook had hoped for—the rules were adopted nevertheless[2]. The global privacy watchdog, Privacy International, called the Facebook vote a "massive confidence trick".[3]

The low voting turnout was largely foreseeable. Obviously, only a small fraction of Facebook users have the time, patience and dedication, and take the service seriously enough to actively participate in its governance. Attempting this kind of *direct democracy* voting in a large social network turned out not to be a good idea. As we discuss in the rest of this section, there exist other voting systems that are more appropriate for online social networks, and that can raise the voting turnout and the credibility of the whole democratic process.

1.1 Transitive Proxy Voting System

A voting system [1, 12] is a set of rules that a community adopts to take collective decisions. It specifies the way the voters express their preference (sometimes called the *ballot*), and the algorithm that determines the final outcome (sometimes called the *tally*). Voting is done for basically one of two purposes: to decide on a *motion* (e.g., to pick the best among a series of alternatives), or to elect a *representative* or set of representatives (e.g., to elect a senate). In both cases the final goal is that of making decisions which reflect as much as possible the opinion of the citizens. The difference is the way in which this is pursued.

The former case, known as *direct democracy*, is based on the idea of ensuring maximum equality and fairness by making all citizens vote directly for the different motions. Direct democracy works better in practice for small cohesive groups. When public decisions reach a certain level of complexity, it becomes impractical for every citizen to become fully informed on every issue.

The latter case, known as *representative democracy*, involves a relatively small number of representatives who are elected by the citizens to take decisions on their behalf. Beyond the issue of which representation structure is the most appropriate for a given context, representative democracy presents certain risks in practice. For instance, by concentrating power in the hands of a small political elite, it creates fertile ground for corruption, entrenchment, conflict of interest, etc., which may result in bad government.

*Part of this work was done while the authors were visiting Yahoo! Research Labs, Barcelona

[1]http://blog.facebook.com/blog.php?post=79146552130

[2]http://www.cio.com/article/490775/After_Vote_Facebook_Gets_New_Governing_Rules

[3]http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-564312

CIKM'09, November 2–6, 2009, Hong Kong, China.

A third way, combining the benefits of direct and representative democracy while avoiding some of their drawbacks, is the so called *delegative democracy* which is based on *proxy voting* [5, 8, 19]. In this context a citizen can decide either to express directly her opinion on an issue, or to delegate her vote to a *proxy*, that is, another citizen that she trusts. If the proxy votes directly on the issue, the weight of all delegated votes she received are added to her vote. Proxy delegation may be transitive: one's vote can be further delegated to her proxy's proxy.

In this novel kind of democracy, also known as *liquid democracy*[4], citizens vote for local candidates who they know and trust personally, rather than distant candidates they know only through reputation, costly mass-media campaigns and televised debates.

This voting system seems well suited for online social networks. In fact, people's social connections can be seen as a mixture of *strong ties* (family, close friends) and *weak ties* (distant friends, acquaintances). Electronically-mediated social networks allow people to maintain many more weak ties than before. This means that the number of connections they have is larger than what one could consider an actual "friendship" network. On the other hand, current social networks are mostly networks of peers, in which the "super-stars" (the individuals with the larger number of connections) do not have the visibility that can be achieved through mass media. The "super-stars" in a social network may be known by only a tiny fraction of the network.

Proxy-voting systems encourage people to cooperate to build direct, permanent political and social relationships with each other and with their individual supporters, forming a web of trust. Everyone can achieve political influence proportional to their level of public support.

1.2 Paper contribution and organization

In this paper we study the problem of voting in electronically-mediated social networks. We focus on *delegative voting systems*, in which votes are transferred through a path of voting choices over the social network, possibly *damping* the vote by an attenuation factor that reduces the strength of the mandate as the vote travels towards farther people. Specifically, we propose a system of *delegable proxy voting with exponential damping*.

Although the theory of voting systems is a well-studied subject in social and political sciences, economics and mathematics, there are some assumptions behind the classical theory that hold true in the traditional scenarios but may fail in electronically-mediated social networks. In particular, we study voting systems in which people can only vote for someone they are explicitly connected to.

In the next section we provide some background and preliminaries for our voting system, bridging some notions of *voting theory* in social science and concepts from graph theory. In this paper, we discuss several design choices in terms of notions from voting theory, translate each choice in the context of the class of voting systems under consideration, and select one alternative for analysis. Due to space constraints, we do not explore in detail every branch of the possible design choices we present in this paper.

In Section 3 we enter in the mathematical characterization of the proposed voting system. We characterize the delegation graph and the possible *winners* and prove several properties of our voting systems.

In Section 4 we study how to deal with missing votes (abstention), including the limit case in which nobody actually expresses a preference. The study of this limit situation leads to the definition of a node centrality measure, dubbed *voting centrality*. We show that in this case the voting score of a node can be seen as a special, simpler case of PageRank for which we provide a closed-form formula.

In Section 5 we tackle some issues that apply specifically to the case of electing multiple representatives. In Section 6 we present experimental results from a simulated vote in two social networks.

[4] http://democracialiquida.es/

2. VOTING IN A SOCIAL NETWORK

The class of social networks under examination are the ones having mutual (symmetric) friendship relationships, that are modeled by means of an *undirected graph* of friendship, say $G = (V_G, E_G)$, defined by a set of vertices (or nodes) V_G and a symmetric set of edges $E_G \subseteq V_G \times V_G$. We let, for every $x \in V_G$, $N_G(x) = \{y \mid (x, y) \in E_G\}$ (the *neighbors* of x in G). For the sake of presentation, we will assume that G is connected; this is not a limiting assumption as our results can be easily extended to deal with disconnected components.

The key aspect of the voting system that we propose for social networks is that votes can be delegated transitively along the existing edges of the social network. That is, any member of the network can choose a proxy among her contacts. A person can also choose not to delegate her vote.

Besides the obvious organizational advantages, the assumption that votes can only be assigned to a direct connection is twofold: on one hand, voters can base their decision on a direct personal knowledge of the person they vote for, making propaganda essentially useless and thus decoupling popularity from credibility; on the other hand, the fact that mandates are attributed through a chain of direct connections should ensure a stronger sense of responsibility.

The tally. In the rest of this paper, unless explicitly stated, we consider voting in general, without distinguishing between voting for motions and voting for choosing representatives, as both can be implemented using the same system. In particular, when voting for motions a weighted count is performed, where each voter that decides not to delegate and instead expresses her preference is counted, weighted by the amount of delegation she has received.

In the case of voting for representatives, a distinction worth considering is that between single-winner and multiple-winner systems. When a committee having s seats must be selected, we can simply select the top-s scoring candidates. However, there are other options, and we defer the discussion of them to Section 5; unless stated otherwise the remainder of the discussion applies to both the single-winner and the multiple-winner case.

The ballot. Various possible choices exist for defining the ballot. *"One-vote"* voting systems are those in which a voter picks exactly one candidate; in our case, one contact. In a *"ranked"* voting system, each voter would rank her contacts in order of preference, and in a *"rated"* voting system, voters would give a score to each contact. In this paper we consider only the *one-vote* kind of ballot, leaving the other two cases for future investigation.

Every node $x \in V$ chooses *exactly one of his neighbors* in G and delegates to her. More formally, the voters' choices are expressed by a *voting function* (or simply "voting"), a function $\upsilon : V_G \to V_G$ such that one has $\upsilon(x) \in N_G(x)$ for all $x \in V$; the set of all voting functions for G is denoted by Vot_G. For every $\upsilon \in \mathrm{Vot}_G$ we let $D(G, \upsilon)$ (the *delegation graph*) be the directed graph (V, A) where $A = \{(x, \upsilon(x)) \mid x \in V\}$. For persons that do not delegate, we include a self-loop in the delegation graph, implicitly assuming a self-loop at each node in the friendship graph.

Proposition 1 *The delegation graph $D = (V, A)$ is a directed, out-regular graph with out-degree 1. D is made of weak components each formed by a cycle with trees (oriented towards their roots) attached to every node in the cycle.*

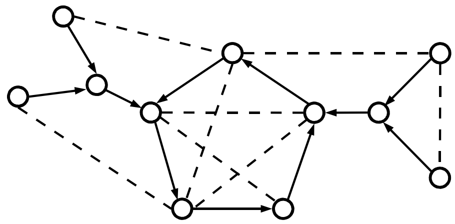

Figure 1: An example of delegation graph over a social network. Delegations are represented by solid arcs, the remainder connections by dashed edges.

Figure 1 illustrates an example of delegation graph.

Interpretation of the ballot. In a delegable proxy voting, given that the vote is delegable, what a node x votes actually affects all the nodes that x transitively votes for—all the nodes along the path to the root or roots of the weakly connected component containing the node. Hence, a "one-vote" delegable proxy voting system can be interpreted as a ballot in which a person (implicitly) votes for several candidates simultaneously. The case of voting for multiple candidates has three main interpretations in voting theory: *cumulative voting* (or weighted voting), *approval voting*, or *preferential voting*.

In a cumulative voting interpretation, $D(G, v)$ describes for each user u an assignment of scores that forms a distribution over the other users, with a weight of 0 for users that can not be reached from u following a delegation path in $D(G, v)$. In an approval voting interpretation, $D(G, v)$ indicates that a user u "approves" all users reachable through a delegation path. These two cases are considered in the exponentially damped system of Section 3.

In a preferential voting interpretation, each user u ranks all the people along the path to a root of the connected components. This adds a level of complexity to the interpretation, not only because of the description of the specific rank aggregation method, but because there is another choice between interpreting the user's ranking as giving more preference to nodes close by than farther away (as in the exponentially damped case), or the exact opposite, which can not be dismissed out of hand. We thus leave the preferential voting interpretation for future work.

3. TRANSITIVE PROXY VOTING SYSTEM WITH EXPONENTIAL DAMPING

In this section, we will present our proposed tally system, the *transitive proxy voting system with exponential damping*. This is similar to standard proxy voting, but with a damping factor that introduces some reluctance in the way delegated votes are transferred. This reluctance corresponds intuitively to the idea that in an electronically mediated social network typically you cannot fully trust your connections, and you want to refrain from giving them all of your delegation. We might call this form of proxy voting a *viscous democracy* (a form of "liquid democracy"), because of the way trust decays with the distance.

The situation is not much different from the typical link-based ranking scheme *à la* PageRank [14], where every vote transfers by transitivity to distances larger than one, but with an attenuation factor; Yamakawa *et al.* [19] use similar ideas, but aim at *total vote estimation* by mixing voters and motions in a single matrix: as a result, their election results are incomparable with ours (see Section 7 for a more detailed comparison).

3.1 Transitive proxy voting and PageRank

In PageRank, the arcs of a directed graph represent hyperlinks between web pages, and a hyperlink is seen as a way to confer authority. More precisely, every time a page x points to a page y, x is transferring a part of its own authority to y; more precisely, if x points to k pages, its authority will be equally distributed among those pages. This process is iterated transitively, with an attenuation factor introduced to limit the risk of forming oligopolies (*rank sinks* in the PageRank jargon).

Although there are many equivalent ways to define PageRank formally, for our purposes it is convenient to introduce it using the path formula of [10]: given a directed graph $D = (V_D, E_D)$, and for a fixed $\alpha \in [0, 1)$ referred to as the *damping factor*, we define the PageRank of node $x \in V_D$ as

$$r_D(x) = \frac{1-\alpha}{n_D} \sum_{\pi \in \mathrm{Path}_D(-,x)} \alpha^{|\pi|} \operatorname{br}(\pi)$$

where $n_D = |V_D|$ (the number of nodes of D), $\mathrm{Path}_D(-, x)$ is the set of all directed paths of D ending in x and the branching of a path is defined as follows:

$$\operatorname{br}(x_1, \ldots, x_{k+1}) = \frac{1}{|N_D^+(x_1)| \ldots |N_D^+(x_k)|},$$

where $N_D^+(z)$ is the set of out-neighbors of z, i.e., $N_D^+(z) = \{w \mid (z, w) \in E_D\}$. Intuitively, every node x receives its importance (its delegations) through incoming paths: every incoming path gives a contribution that depends on the number of branches the path contains, and that decays exponentially with its length. The PageRank formula can also be applied to an *undirected* graph, by looking at an edge $e = \{x, y\}$ as if it were a pair of arcs (x, y), (y, x).

Let us adopt the PageRank method in our context: we define the *(voting) score* of node x for the voting function v as $r_{D(G,v)}(x)$, the PageRank of x in the directed graph $D(G, v)$. Note, however, that the PageRank formula applied to such simple graphs turns out to be much easier to analyze. In a graph with out-degree 1, the branching factor is always 1; so we can write:

$$r_i = \frac{(1-\alpha)}{n} \sum_{p \in \mathrm{Path}(-,i)} \alpha^{|p|}.$$

This voting system depends on a single parameter α. In the following, we discuss its properties for "small enough" and "large enough" values of α. It is worth noting that for any reasonable value of α the ballot has a cumulative interpretation, while the approval interpretation corresponds to the voting system with no damping factor (i.e., when $\alpha \to 1$).

3.2 Criteria that hold for this system

In this section, we will collect a number of properties that hold true for this voting system. For sake of presentation, we assume to vote for selecting representatives, i.e, the final result of the election is a non-empty set of *winners*, that depends on the votes cast by the nodes and on a parameter $\alpha \in [0, 1)$ that determines how "transitive" a vote is ($\alpha = 0$ means that the vote is not transitive at all, i.e., that it does not propagate). A node x is *transient* if it votes for a node that does not transitively vote for x (i.e., iff it is a non-root node of a tree), *non-transient* otherwise.

Undelegable mandates. The intuitive desire that a strong mandate cannot be delegated is satisfied at some extent in our voting system, as the following proposition shows:

Proposition 2 *Suppose that x votes for y, y votes for z, and y, z do not receive other votes. Then the score of y is larger than the score of z iff the score of x is larger than the average.*

PROOF. The ranks of y and z expressed in terms of the rank of x are

$$r_y = \frac{1-\alpha}{n}\left(1+\alpha\frac{n}{1-\alpha}r_x\right)$$
$$r_z = \frac{1-\alpha}{n}\left(1+\alpha+\alpha^2\frac{n}{1-\alpha}r_x\right).$$

Hence, the condition $r_y > r_z$ can be written as

$$\frac{n}{1-\alpha}r_x > 1+\alpha\frac{n}{1-\alpha}r_x$$

which is true iff $r_x > 1/n$, the average score. □

In other words, if y receives a mandate that is greater than the average, she will obtain a score that is greater than the one she can delegate; the average score is a threshold over which a received mandate cannot be completely delegated.

Large damping (small α) produces undelegable mandates. Consider the delegation graph among $k+2$ members shown in Figure 2, and assume that x votes for herself.

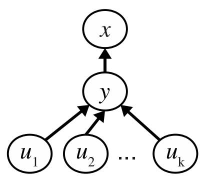

Figure 2: If the damping factor is large enough, y will be the winner, otherwise x will be the winner.

Intuitively, we should consider that if k is large enough, then member y is receiving a strong mandate. Should then y represent the community, independently from her choice of delegating to member x? Using the same idea as in Proposition 2, y wins over x iff $(1-\alpha)k/n > 1/n$, that is $\alpha < 1/2 - 1/k$ (because there is too much attenuation for the delegation to be transferred to x); of course, larger values of α will make x win, as she would in a traditional proxy system.

This can be easily explained to voters using this voting system. The choice of the damping factor can be described as the choice of the critical number of voters k^* that creates an "undelegable" mandate in this situation (of course other voters can affect the outcome, but y alone can not). More precisely, if we want that k^* or more voters always make the mandate to y undelegable, we have to choose α not larger than $1/2 - 1/k^*$.

Large damping (small α) makes only direct votes count. We prove the following:

Proposition 3 *For sufficiently small α, all winners have the same maximum number of voters (i.e., the same indegree in the delegation graph).*

PROOF. Suppose that x is has a voters, y has b voters and $a > b$. Then the score of x is $1 + a\alpha + \alpha^2 A(\alpha)$ whereas the score of y is $1 + b\alpha + \alpha^2 B(\alpha)$, for suitable polynomials A and B with positive coefficients. The score of x is smaller than the score of y iff

$$1 + a\alpha + \alpha^2 A(\alpha) < 1 + b\alpha + \alpha^2 B(\alpha)$$

which implies

$$(a-b)\alpha < \alpha^2(B(\alpha) - A(\alpha))$$
$$\frac{a-b}{\alpha} < B(\alpha) - A(\alpha)$$

which can never be true when α is sufficiently close to zero (the left-hand side goes to $+\infty$, whereas the right-hand side is finite). □

As a corollary, we have:

Corollary 1 *For sufficiently small α, the vote of every node x can only influence the victory of its neighbors (i.e., it can add or remove only one of the neighbors of x from the set of winners).*

Small damping (large α) determines winners by size of subtrees. Clearly, at least when α is large enough, the nodes that obtain the largest rank are the roots of all trees (we shall prove this formally in Section 3.2): suppose that a node is the root of a tree T of height $d(T)$, with $n_0(T) = 1$ nodes at depth 0 (just the root), $n_1(T)$ nodes at depth 1 (the root's children), $n_2(T)$ nodes at depth 2 etc. Then the rank of the root node $\rho(T)$ is proportional to

$$\rho(T) = \sum_{t=0}^{d(T)-1} n_t(T)\alpha^t.$$

At this point, the leader will be chosen among the nodes that participate in the cycles (the roots of trees); suppose that a cycle contains k nodes, say $x_0, \ldots, x_{k-1}$, where x_i is the root of a tree T_i. Then the rank of node i is a combination of the $\rho(T_j)$ obtained by:

$$r_i = \frac{(1-\alpha)}{n}\sum_{t=0}^{k-1}\rho(T_{i-t})\frac{\alpha^t}{1-\alpha^k}.$$

In the voting system under discussion, we can describe exactly what happens in the case $\alpha \to 1$, that is, when no damping is introduced and the votes transfer transitively. We have that $\rho(T)$ ends up being proportional to the number of nodes of T. Moreover, we note that

$$r_i \propto \sum_{t=0}^{k-1}\rho(T_{i-t})\frac{\alpha^t(1-\alpha)}{1-\alpha^k} =$$
$$= \sum_{t=0}^{k-1}\rho(T_{i-t})\frac{\alpha^t}{1+\alpha+\alpha^2+\cdots+\alpha^{k-1}} \to$$
$$\to \frac{1}{k}\sum_{t=0}^{k-1}\rho(T_{i-t}) \qquad \text{as } \alpha \to 1.$$

Thus, when $\alpha \to 1$ the winners are (*ex-aequo*) the nodes on the cycle belonging to the component with the largest average tree size. In the voting theory jargon this set of nodes is called *minimal dominant set*. When in a voting system there is a minimal dominant set of winners, the method is said to be *Condorcet-efficient* [8]. If no cycle exists, the winner (i.e., the top-scorer) is precisely the root of the largest tree (or the roots of the largest trees, if more than one tree of largest size exists), as prescribed by proxy voting.

Small damping (large α) makes transient nodes lose. We prove the following:

Proposition 4 *For sufficiently large α, all winners are non-transient.*

PROOF. Suppose that x is transient and votes for y. The score of x is $(1-\alpha)S(\alpha)/n$, where S is a polynomial with positive coefficients (the coefficient of α^k being the number of descendants of x at depth k), whereas the score of y is at least $(1-\alpha)(1+\alpha S(\alpha))/n$. The score of x is smaller than the score of y iff

$$S(\alpha) < 1 + \alpha S(\alpha);$$

if t is the sum of all the coefficients of S, the left-hand side tends to t as $\alpha \to 1$, whereas the right-hand side tends to $t+1$, so the above inequality is ultimately true, hence x cannot be a winner for sufficiently large α. □

Bolzano-Weierstrass's theorem for voting. It is worth observing that the scores produced by our voting system change continuously with α:

Proposition 5 *If $\alpha_0 \le \alpha_1$, x is a winner at α_0 and y is a winner at α_1, then there is some $\alpha \in [\alpha_0, \alpha_1]$ such that both x and y are winners at α.*

PROOF. Since the score r is a rational function of α with no poles in [0, 1], the statement is trivial by continuity. □

Ultimate sovereignty. The condition that every node can become a winner is known as *surjectivity* or *sovereignty*.

Proposition 6 *If the graph G is reflexive (i.e., if everybody can vote for himself), for sufficiently large α every node can become a winner.*

PROOF. Due to Proposition 4, it is sufficient to prove that for all node x there is a $v \in \mathrm{Vot}_G$ such that x is the only non-transient in $D(G, v)$. Consider a spanning tree for G (recall that G is connected), and root it at x; then, make every node vote for his parent in the tree, and let x vote for itself. □

The reflexivity condition is needed, as explained in the following counterexample: consider a graph with $n-2$ nodes $z_1, \ldots, z_{n-2}$ attached to a node y, and the latter adjacent to the node x. In every voting, z_i can only vote for y, and so does x; y can choose whether to vote for x or for some z_i, but if we want x to be a winner, y must vote for x. In this condition, though, the score of x and y are

$$\begin{aligned} r_x &= \frac{1+\alpha+(n-2)\alpha^2}{n(1+\alpha)} \\ r_y &= \frac{1+(n-1)\alpha}{n(1+\alpha)} \end{aligned}$$

and r_x is always smaller than r_y, so x can never be a winner.

4. DEALING WITH MISSING VOTES

In a realistic scenario, only a (probably small) fraction of users actually votes: recall the case discussed in the introduction. We would like to consider how the election can still be carried out in the presence of abstentionism.

4.1 Voting without voting

We start by studying the limit situation in which nobody actually expresses a preference. Our approach is to compute the expected result of the election if voters were to decide their vote uniformly at random, of course subjected to the constraint of voting for someone they are explicitly connected to. As we show in this section, this induces a novel measure of centrality for nodes on a graph.

Formally, we define the *voting centrality* of a node x as follows:

$$\hat{r}_G(x) = \frac{1}{|\mathrm{Vot}_G|} \sum_{v \in \mathrm{Vot}_G} r_{D(G,v)}(x) ,$$

which can be more explicitly rewritten as

$$\hat{r}_G(x) = \frac{1-\alpha}{n_G |\mathrm{Vot}_G|} \sum_{v \in \mathrm{Vot}_G} \sum_{\pi \in \mathrm{Path}_{D(G,v)}(-,x)} \alpha^{|\pi|} .$$

In the rest of this section we produce a closed formula for this measure. We start by observing that for a voting function v, a path $\pi \in \mathrm{Path}_{D(G,v)}(-, x)$ is either a simple path of the friendship graph G ending in x, or a simple path with a simple cycle containing x attached at the end, possibly repeated many times: no other case is possible, because $D(G, v)$ has out-degree 1 so it can only contain a cycle at the end, if any.

Definition 1 *Given two paths $\pi_1, \pi_2 \in \mathrm{Path}_G(-, x)$, we write $\pi_1 \perp_{G,x} \pi_2$ to mean that the following conditions hold:*

- *π_1 is a simple path;*
- *π_2 is a simple path or a simple cycle;*
- *the starting node (say y) of π_2 appears in π_1;*
- *except for x and y, π_1 and π_2 are disjoint.*

Under these conditions, we let $\pi_1 - \pi_2$ be the part of the path π_1 that appears after y, and we define the set

$$S(\pi_1, \pi_2) = \{\pi_1(\pi_2^R(\pi_1 - \pi_2))^k, k > 0\}.$$

Note that this set either contains only π_1 (if $|\pi_2| = 0$), or it is infinite.

This definition is illustrated in Fig. 3, where we show two paths π_1 and π_2 and how they are oriented in $S(\pi_1, \pi_2)$.

We have that:

Proposition 7 *Let $\pi_1 \perp_{G,x} \pi_2$. Then:*

1. *$S(\pi_1, \pi_2)$ are disjoint for different pairs (π_1, π_2);*
2. *for every $\rho \in S(\pi_1, \pi_2)$, there are exactly*
$$|\mathrm{Vot}(G)|\, \mathrm{br}(\pi_1 \pi_2^R)$$
elements $v \in \mathrm{Vot}(G)$ such that $\rho \in \mathrm{Path}_{D(G,v)}(-, x)$;
3. *if $\rho \in \mathrm{Path}_{D(G,v)}(-, x)$ for some v, then there are $\pi_1' \perp_{G,x} \pi_2'$ such that $\rho \in S(\pi_1', \pi_2')$.*

PROOF. From $\rho \in S(\pi_1, \pi_2)$ one can deduce π_1 and π_2 uniquely as follows: π_1 is the first part of ρ until the first occurrence of x; π_2^R is the following part of the path until the first repeated node (it will be a cycle if the first node repeated in ρ is x itself). This proves (1) and (3). To prove (2), notice that, for $\pi_1 = (x_1, \ldots, x_k)$ and $\pi_2 = (y_1, \ldots, y_h)$ you must define $v(x_i) = x_{i+1}$ for all $0 < i < k$, and $v(y_{i+1}) = y_i$ for all $0 < i < h$; $v(-)$ can be freely defined on all the remaining nodes. □

The lengths of the paths in $S(\pi_1, \pi_2)$ are $|\pi_1| + k(\ell + |\pi_2|)$ where $\ell = |\pi_1 - \pi_2|$ and $k > 0$. Now

$$\sum_{k=1}^{\infty} \alpha^{|\pi_1| + k(\ell + |\pi_2|)} = \frac{\alpha^{\ell + |\pi_1| + |\pi_2|}}{1 - \alpha^{\ell + |\pi_2|}} .$$

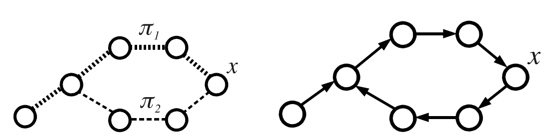

Figure 3: Two undirected paths $\pi_1 \perp_{G,x} \pi_2$ in the friendship graph, and with their orientation in $S(\pi_1, \pi_2)$.

where we understand that the right-hand side should evaluate to $\alpha^{|\pi_1|}$ if $|\pi_2| = 0$.

As a consequence, we have the following explicit closed formula for voting centrality:

Theorem 1 *The voting centrality of a node x in a graph G is given by*

$$\hat{r}_G(x) = \frac{1-\alpha}{n_G} \sum_{\pi_1 \perp_{G,x} \pi_2} \frac{\alpha^{|\pi_1-\pi_2|+|\pi_1|+|\pi_2|}}{1-\alpha^{|\pi_1-\pi_2|+|\pi_2|}} \mathrm{br}(\pi_1 \pi_2^R). \quad (1)$$

4.2 Partial abstention

After studying the limit case of total abstention, we return to the more realistic case of partial abstention. We deal with this case by considering a partially directed graph that contains both directed and undirected edges. Essentially we reduce to the previous case of total abstention, considering the real votes expressed as sort of constraints. In practice, Equation 1 still holds provided that the simple paths π_1 and π_2^R are chosen respecting the direction on the directed arcs.

We remark that Equation 1 expresses the voting centrality as a *finite* summation (there are only a finite number of pairs $\pi_1 \perp_{G,x} \pi_2$), and enumerating the pairs diagonally allows one to stop the evaluation at a desired degree of precision.

4.3 A random surfer interpretation of voting centrality

Beyond its interpretation as voting score in the case of "total abstentionism", voting centrality is a new measure of centrality in a graph that is worth further analysis.

In the following we provide a random-surfer interpretation of voting centrality, but we shall state it in a more general way, as an expected PageRank on a given graph distribution.

Expected PageRank: the general case. Let V be a fixed set of n elements and $\mathscr{G}$ be the set of all directed graphs with node set V.

Let $P(-)$ be a probability distribution on $\mathscr{G}$, and for each $G \in \mathscr{G}$ let $\boldsymbol{v}_G$ be a distribution on its nodes. Let $\boldsymbol{r}_G$ be the PageRank vector for the graph $G \in \mathscr{G}$ using the preference vector $\boldsymbol{v}_G$, that is,

$$\boldsymbol{r}_G = (1-\alpha)\boldsymbol{v}_G(I - \alpha A_G)^{-1}$$

where A_G indicates the row-normalized adjacency matrix[5] of G.

We are interested in the expected PageRank vector $\hat{\boldsymbol{r}} = E[\boldsymbol{r}_G]$, that is

$$\hat{\boldsymbol{r}} = \sum_{G \in \mathscr{G}} P(G)\boldsymbol{r}_G.$$

Consider a random surfer acting as follows. At each time, the surfer is in some node of some graph of $\mathscr{G}$. When the surfer is in the node x of a graph G, with probability α, he moves to one of the out-neighbors of x at random; with probability $1-\alpha$ he resets, i.e., he chooses a graph G (according to the probability distribution $P(-)$) and a node in G, according to $\boldsymbol{v}_G$, and moves there.

Equivalently, one can think of the surfer as surfing on a single large graph H that is the disjoint union of all the graphs in $\mathscr{G}$. For the sake of simplicity, let $V = \{1, \ldots, n\}$ and $\mathscr{G} = \{G_1, \ldots, G_k\}$, and let us write $\boldsymbol{v}_i$ for $\boldsymbol{v}_{G_i}$ and P_i for $P(G_i)$. Then H will have nk nodes, and its row-normalized adjacency matrix will be as follows:

$$A_H = \begin{pmatrix} A_{G_1} & 0 & 0 & \ldots & 0 \\ 0 & A_{G_2} & 0 & \ldots & 0 \\ 0 & 0 & A_{G_3} & \ldots & 0 \\ & & \ldots & \ldots & \\ 0 & 0 & 0 & \ldots & A_{G_k} \end{pmatrix}$$

where 0 denotes an $n \times n$ matrix of zeros. The vector

$$\boldsymbol{v} = (P_1\boldsymbol{v}_1, P_2\boldsymbol{v}_2, \ldots, P_k\boldsymbol{v}_k)$$

describes how the surfer chooses the pair graph/node when he decides to reset. So, the surfer's behavior is described by the transition matrix:

$$\alpha A_H + (1-\alpha)\mathbf{1}^T\boldsymbol{v},$$

whose stationary distribution is given by

$$(1-\alpha)\boldsymbol{v}(I - \alpha A_H)^{-1}.$$

Now $(I - \alpha P_H)^{-1}$ can be written as

$$\begin{pmatrix} I - \alpha A_{G_1} & 0 & \ldots & 0 \\ 0 & I - \alpha A_{G_2} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ & & \ldots & \ldots \\ 0 & 0 & \ldots & I - \alpha A_{G_k} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} (I - \alpha A_{G_1})^{-1} & 0 & \ldots & 0 \\ 0 & (I - \alpha A_{G_2})^{-1} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ & & \ldots & \ldots \\ 0 & 0 & \ldots & (I - \alpha A_{G_k})^{-1} \end{pmatrix}$$

so $(1-\alpha)\boldsymbol{v}(I - \alpha A_H)^{-1}$ is equal to

$$(1-\alpha)(P_1\boldsymbol{v}_1(I - \alpha A_{G_1})^{-1}, P_2\boldsymbol{v}_2(I - \alpha A_{G_2})^{-1}, \ldots, P_k\boldsymbol{v}_k(I - \alpha A_{G_k})^{-1}) = (P_1\boldsymbol{r}_1, \ldots, P_k\boldsymbol{r}_k).$$

Now, given a node $x \in V$, the stationary probability that the surfer is in x is the probability that he is in one of the k copies of x, that is, $P_1(\boldsymbol{r}_1)_x + \cdots + P_k(\boldsymbol{r}_k)_x = \hat{\boldsymbol{r}}_x$.

Expected PageRank: the case of voting. In the case of voting, we can state the surfing activity on the undirected graph G as follows. Whenever the surfer is in some node x:

- with probability α,
  - if x was never visited since the last reset, he moves to one of the out-neighbors of x at random, and stores this information in his own memory;
  - if x was already visited, he moves to the same neighbor where he moved the last time he visited x;
- with probability $1-\alpha$ it resets, i.e., moves to a random node and cleans its memory.

It is easy to see that this is just an equivalent restatement of our previous general surfer, hence the average time spent on x will be precisely the voting centrality of x.

5. SELECTING MULTIPLE WINNERS

In the case of an election for electing several representatives, there is an opportunity of selecting a committee that represents the diversity of users, by ensuring *proportionality*. The criterion of proportionality states that each group should have a share of the seats roughly equal to its share of the votes [12].

[5] Null rows are patched in the usual way, substituting them by rows full of $1/n$.

	C_1	C_2	C_3
1^{st} seat	**12**	8	5
2^{nd} seat	6	**8**	5
3^{rd} seat	**6**	4	5

(a)

	C_1	$C_2 \cup C_3$
1^{st} seat	12	**13**
2^{nd} seat	**12**	6.5
3^{rd} seat	6	**6.5**

(b)

Figure 4: Example of proportional voting using D'Hondt rules and three seats. Numbers in boldface represent elected candidates.

In most existing voting systems, the groups used as a base for measuring the proportionality of a committee are (i) voting districts; (ii) political parties or alliances; or (iii) a mixture of both. The concept of *voting districts* can be mapped easily into voting systems for on-line social networks. For instance, the voting can use explicit non-overlapping groups to which users belong (in the case of social networks that support such groups). If these explicit groups are not available, then as social networks exhibit community structure [6], the friendship graph G can be partitioned using a graph clustering method and then members selected proportionally to the sizes of each cluster. As in real-world voting districts, this allocation must be done *before* the voting takes place, and users should be able to vote only for their friends belonging to the same district as themselves. Furthermore, the allocation can be done according to the structure of the network. A well connected network may need fewer representatives and fewer districts than a more disconnected network.

The concept of *parties or alliances* can also be mapped easily into voting systems for on-line social networks. In this case, the allocation of users into parties is done *after* the voting takes place, as it depends on the voting. When no explicit alliances are declared beforehand, a voting system for social networks may interpret the connected components of the delegation graph as alliances, as they represent communities of like-minded people who delegate to other members of their community but not to external people.

5.1 Proportionality using the delegation graph

Let us examine the case in which each connected component of $D(G, v)$ is considered an alliance, and we want to ensure alliances are represented proportionally. The number of nodes in each component corresponds to the number of people that voted for the corresponding alliance. Let us assume in $D(G, v)$ there are k weakly connected components $C_1, C_2, \dots C_k$ with sizes $m_1, m_2, \dots m_k$. Each alliance is considered an *open list* of candidates, in the sense that their ordering is not fixed in advance, but determined by the popular vote, e.g. by using the obtained scores.

The method for allocating the number of seats for each alliance can be determined by any system for proportional voting. For concreteness, we describe here the use of D'Hondt rules [12], which are a proportional representation system used for the parliament of several countries, and in the European Parliament elections. Nevertheless we stress that most of the arguments that follow apply to other proportional voting systems.

Under D'Hondt rules seat allocation is done round-wise. In each round the new seat is assigned to the party with the highest ratio $m/(s+1)$, where m is the number of votes received, and s is the number of seats that party has been allocated so far. An example tally is shown in Figure 4(a). In the example $m_1 = 12, m_2 = 8, m_3 = 5$ and $s = 3$, the numbers in the table are the ratios, and the first group gets two seats, the second group gets one seat, and the last group gets zero seats.

5.2 Strategic voting due to non-monotonicity

Monotonicity is another desirable property of voting systems[6]. It states that if an alternative x loses, and the ballots are changed to disfavor x, then x must still lose. As in many other multiple-winner voting systems in our voting system there is *no monotonicity*[7]. As a counterexample, let us consider the top scoring node t_3 of component C_3 in Fig. 4(a), and let us assume that there are no cycles in the delegation graph. It would have been better for this node to vote for the top node t_2 of C_2, to create a larger community $C_2 \cup C_3$ that can elect two representatives under these rules. In some circumstances (e.g. if C_2 were star-shaped, so that t_2 is the only one having in-degree larger than zero in C_2) then t_3 would be the 2-nd best node in the component $C_2 \cup C_3$ and it would have been elected, as shown in Fig. 4(b).

The non-monotonicity is not limited to the case of non-delegation (or self-delegation). A non-elected node x in community C_i, by defecting to another community, e.g., voting for an elected node y in community C_j, might under certain circumstances (i.e. given a large enough damping factor), obtain a score larger than y thus be elected, possibly displacing y from the committee if the target community does not have enough seats for both.

In general, in the case of strategic voting, as the vote of x transitively affects all the nodes in the path to the root of the community of the user receiving her vote, it is in the best interest of x to vote for someone as close to the root as possible. This minimizes her contribution to the score of other nodes and thus increases the likelihood of being among the top members in her community.

6. EXPERIMENTS

The effectiveness of a voting system in practice depends on qualitative factors, such as whether the decisions reached by the community are in some sense correct and whether the members of the community accept such decisions. In this experimental section, we focus instead on measurable aspects of the exponentially damped voting system, such as its stability with respect to the choice of a damping factor, and its correlation with simpler measurements on social networks.

For these experiments, we picked two different social networks, one of scientists (the DBLP co-authorship network) and one of photographers (the Flickr social network).

6.1 The DBLP collaboration network

DBLP[8] is a bibliography service from which a scientific collaboration network can be extracted. In this network each node represents a scientist and we interpret the co-authorship relation as a social tie, indicating that two scientists are connected if they have worked together on an article. The strength of the connection is the number of articles they have co-authored. We refer to this graph as the *DBLP social network*: as of now, it contains 326 186 scientists and 1 615 400 co-authorship relations. This particular social network has been studied in the past e.g. in [4], and it has been shown that it shares many properties of a typical social network, including the small-world property.

We simulated a voting process in the DBLP social network, in which the community elects a committee of top scientists. This is done by each person delegating his/her vote in one of his/her

[6]Some political scientists, however, doubt the value of monotonicity as a property of voting systems: *"monotonicity in electoral systems is a non-issue: depending on the behavioral model governing individual decision making, either everything is monotonic or nothing is monotonic."* [2]

[7]*Arrow's impossibility theorem*, or *Arrow's paradox*, demonstrates that no voting system can have all the desirable properties [1].

[8]http://www.informatik.uni-trier.de/~ley/db/

	DEG	PR	PROD	CEN	r^A	r^B	r^C	$r^{D,0.25}$	$r^{D,0.50}$	$r^{D,0.75}$
DEG	1.00	0.61	0.92	0.33	0.12	0.17	0.19	0.25	0.25	0.25
PR	0.61	1.00	0.58	0.67	0.46	0.48	0.53	0.49	0.46	0.44
PROD	0.92	0.58	1.00	0.31	0.12	0.20	0.19	0.24	0.24	0.24
CEN	0.33	0.67	0.31	1.00	0.46	0.47	0.53	0.43	0.40	0.37
r^A	0.12	0.46	0.12	0.46	1.00	0.84	0.75	0.52	0.48	0.44
r^B	0.17	0.48	0.20	0.47	0.84	1.00	0.69	0.50	0.46	0.43
r^C	0.19	0.53	0.19	0.53	0.75	0.69	1.00	0.54	0.51	0.46
$r^{D,0.25}$	0.25	0.49	0.24	0.43	0.52	0.50	0.54	1.00	0.42	0.39
$r^{D,0.50}$	0.25	0.46	0.24	0.40	0.48	0.46	0.51	0.42	1.00	0.38
$r^{D,0.75}$	0.25	0.44	0.24	0.37	0.44	0.43	0.46	0.39	0.38	1.00

(a)

	DEG	PR	PROD	CEN	r^A	r^B	r^C	$r^{D,0.25}$	$r^{D,0.50}$	$r^{D,0.75}$
DEG	1.00	0.43	0.38	0.26	0.30	0.20	0.25	0.06	0.07	0.06
PR	0.43	1.00	0.26	0.60	0.29	0.21	0.29	0.08	0.09	0.06
PROD	0.38	0.26	1.00	0.18	0.27	0.21	0.25	0.06	0.06	0.05
CEN	0.26	0.60	0.18	1.00	0.20	0.19	0.27	0.06	0.07	0.05
r^A	0.30	0.29	0.27	0.20	1.00	0.22	0.37	0.04	0.09	0.05
r^B	0.20	0.21	0.21	0.19	0.22	1.00	0.21	0.08	0.05	0.03
r^C	0.25	0.29	0.25	0.27	0.37	0.21	1.00	0.06	0.06	0.06
$r^{D,0.25}$	0.06	0.08	0.06	0.06	0.04	0.08	0.06	1.00	0.02	0.01
$r^{D,0.50}$	0.07	0.09	0.06	0.07	0.09	0.05	0.06	0.02	1.00	0.01
$r^{D,0.75}$	0.06	0.06	0.05	0.05	0.05	0.03	0.06	0.01	0.01	1.00

(b)

Figure 5: DBLP dataset: (a) Kendall's τ, and (b) Jaccard coefficient of the top-100 set.

connections (co-authors). We will assume there are two factors that influence how this delegation is done: productivity (we use number of papers written as a proxy for productivity), and strength of the co-authorship relation (i.e., number of papers co-authored).

More precisely, we ran a number of elections that all use the above criteria in different combinations:

- **Election A** (*productivity only*). x votes for the most prolific among her co-authors; ties are broken using strength of co-authorship.
- **Election B** (*strength of co-authorship only*). x votes for her main co-author (the one with which x wrote most papers); ties are broken using productivity.
- **Election C** (*adjusted co-authorship*). x considers her co-authors in decreasing order of strength of co-authorship, and votes for the first one that has been more productive than x (i.e., that has written more papers than x); if no such co-author exists, x votes for herself.
- **Election D** (*mixture*). In this case, we use a probabilistic procedure; a parameter $\gamma \in [0, 1]$ is fixed, and for every author x we establish a distribution among the co-authors of x, that is the convex combination of normalized co-authorship strength (with weight γ) and normalized productivity (with weight $1-\gamma$). Then, x votes for one of her co-authors chosen according to the distribution.

We performed six elections (D was run three times, with $\gamma = 0.25, 0.5, 0.75$, respectively), and obtained six score vectors that we shall denote with r^A, r^B, r^C, $r^{D,0.25}$, $r^{D,0.50}$, $r^{D,0.75}$. Here, and in the following, unless otherwise specified we used the standard damping factor $\alpha = 0.85$.

Comparison with link-based metrics. Our first set of experiments aims at determining how much the result of the election depends on the voters' choice, and whether and how it can be determined by the underlying social network instead. For this purpose, we computed four more *static* scorings based on the DBLP social network: the degree DEG (where the score is the number of co-authors), the PageRank PR, the productivity PROD (where the score is the number of papers written), and the voting centrality CEN introduced in Sec. 4. We compared the ten scoring vectors using Kendall's τ and the Jaccard coefficient of the top 10, 100 and 1 000 authors: since these measures substantially agree, we report only the top 100 case in Fig. 5.

Although all scores are positively correlated, they markedly differ. The static measures are more highly correlated among them (in particular, as expected, degree and PageRank), but they are not well correlated with the results of voting, except very mildly for PageRank. On the other hand, the results of the six elections are different, and their Kendall's τ never exceeds 0.84, and it is often around 0.50; as for the Jaccard coefficient, it is at most 0.37, and often under 0.1.

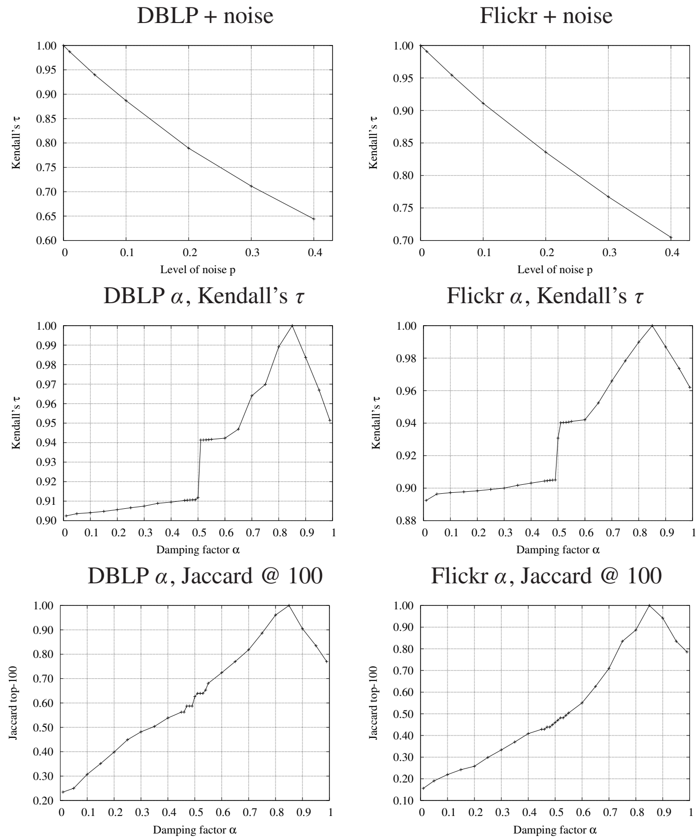

Figure 6: Stability of the voting results in DBLP (left) and Flickr (right). Stability is studied with respect to: noisy preferences, changes in α reflected by Kendall's τ, and changes in α measured by Jaccard at 100.

We can conclude that the outcome of the voting system strongly depends on the voters' choices and not only (or too strongly) on the graph structure or on the position of the node in the graph, both being desirable properties in a democratic voting system for an online social-network.

Stability with respect to preferences. There is a second, somewhat dual question about the stability of the proposed voting system with respect to small changes in the preferences of the community. How does a small change in the voters' choice affect the outcome of the election? To answer this question, we randomized the vote according to a noise parameter. With probability p, a voter simply votes for one of her connections at random, and with probability $1 - p$ she makes her choice as established. We then compare the resulting score vectors with the one obtained without noise.

	Deg	Deg$^+$	PR	Faves	Cen	r
Deg	1.00	0.88	0.81	0.56	0.66	-0.04
Deg$^+$	0.88	1.00	0.77	0.50	0.64	-0.06
PR	0.81	0.77	1.00	0.54	0.83	0.02
Faves	0.56	0.50	0.54	1.00	0.50	0.08
Cen	0.66	0.64	0.83	0.50	1.00	0.07
r	-0.04	-0.06	0.02	0.08	0.07	1.00

(a)

	Deg	Deg$^+$	PR	Faves	Cen	r
Deg	1.00	0.59	0.60	0.16	0.40	0.08
Deg$^+$	0.59	1.00	0.43	0.04	0.29	0.02
PR	0.60	0.43	1.00	0.17	0.60	0.08
Faves	0.16	0.04	0.17	1.00	0.13	0.16
Cen	0.40	0.29	0.60	0.13	1.00	0.06
r	0.08	0.02	0.08	0.16	0.06	1.00

(b)

Figure 7: Flickr dataset: (a) Kendall's τ, and (b) Jaccard coefficient of the top-100 set.

In Fig. 6 (top left), we present the results for the election C above. As the graph shows, even altering 10% of the votes leaves the scores substantially unaltered (τ is above 0.90). Beyond that, the decay is very smooth, apparently linear, and there is a good positive correlation even at high noise levels, proving that the system is stable and reasonably robust against noise.

Stability with respect to the damping factor. Finally, we consider how the score depends on the damping factor. For a fixed election, we computed the voting scores for different values of α and compared them with the voting score obtained with $\alpha = 0.85$. The comparison was performed using Kendall's τ and looking at the Jaccard coefficient at 10, 100 and 1 000. In Figure 6 (middle and bottom) we show Kendall's τ and Jaccard at 100 for election C above. As expected there is a dependency, and the results change smoothly as α moves away from its starting value of 0.85. In particular, Kendall's τ is always above 0.95 for all values of α larger than 0.65 (Jaccard is over 75% in the same region). It is curious to observe that τ has an abrupt change around $\alpha = 0.5$; interestingly enough, this happens on all the examples we considered, and more generally on all tree-shaped directed graphs. Our conjecture is that this phase transition is related to the undelegable mandates described on Sec. 3.1, the phenomenon still requires further analysis.

6.2 The Flickr social network

Flickr[9] is an online community were users can share photographs and videos. The interaction mechanism in Flickr is rich. For instance, users can comment on each other photos and maintain a collection of their *favorite* photos. In Flickr the notion of acquaintance is modeled through *contacts*. Differently from other social networks, in Flickr contacts requests do not need to be confirmed and are thus directed, but we used them disregarding their direction for the purpose of building the friendship graph. We sampled 25 million users from Flickr and restricted our attention to the largest connected component of contacts (526 606 users).

We simulated a voting process on this friendship network as follows: every user x votes for his contact y that has the most photographs selected by x as "favorites" (with ties broken arbitrarily). This voting process produces a voting score vector r.

As for the DBLP graph, for comparison we considered other metrics: the number of incoming or outgoing contacts Deg, the number of incoming contacts Deg$^+$, the PageRank PR (computed on the contact graph), the voting centrality Cen and the overall users' favorites Faves. The latter contains, for every user, the number of times one of her pictures was selected as a "favorite" by anyone in the network (not necessarily by one of her contacts).

The correlation between such measures are shown in Fig. 7, and once more we observe that the results of the voting are not determined by any of the static metrics. Also the measures of stability shown in Fig. 6 (right) exhibit the same behavior as in the DBLP dataset.

Finally, we can look at a more fine-grained measure on the top 100 users: the number Faves of their photos selected as "favorites" by other users, weighted by the scores assigned by them by the different ranking functions. We normalized the scores so that they add up to 1 when considering the top 100 users. The results, in order of similarity to Faves, are $r = 27$K, PR$= 22$K, Deg$= 21$K, Cen$= 18$K, and Deg$^+ = 8$K. The results of the voting outperform the other metrics when compared with Faves.

7. RELATED WORK

Before the conclusions section, we summarize here the previous work in link-based methods for aggregating the preferences of individuals.

Ranking by some kind of eigenvalue-related technique had its origin with the work of Seeley [16]. In modern language, given a square matrix M expressing preferences between individuals, Seeley proposes to compute the dominant eigenvector of M after row normalization (in other words, PageRank with $\alpha = 1$). Independently, a few years later Katz [11] considered for the same purpose the power sum $\mathbf{1}M \sum \alpha^n M^n$ (in other words, PageRank without row normalization—see [3]) for suitable values of α. Many variants of these two basic ideas have appeared in the following literature, and by now *spectral ranking* is a standard tool in many fields. It should be noted, however, that the same technique is used to estimate different concepts such as *authority*, *power*, *influence* and *centrality*.

A spectral-ranking technique for delegated voting was proposed by Yamakawa *et al.* [19]. Assuming not all voters might want to choose an option out of a given set, they propose computing PageRank (therein called the *Bonacich index*) on a stochastic matrix containing both voters and motion. The results of this method are incomparable to ours: if motions are just a copy of the voters, the resulting graph has a completely different structure, as at least half of the nodes will be without successors. In case nobody votes, they propose to use the scores obtained (which are now identical to our ranks) to select powerful voters and force them to express their opinion: however, at each selection the matrix is altered by deleting all delegations of the selected voter, so their order in choosing representatives does not coincide with ours. Finally, they do not consider the interplay between an underlying social network of acquaintance and the votes, which for example would make it impossible to account for missing votes (in their terminology, missing delegation).

Reference [18] deals with a decentralized voting system for object reputation, but there is no collective decision, that is, every peer decides how much to trust on the votes of others.

Our work is also related to the general issue of developing decision-support systems for social networks [15], to the study of trust [9, 7] and influence propagation [13] in social networks.

8. CONCLUSIONS

In this paper we wore hat and glasses: we wore the political scientist's hat by concretely proposing a practical voting system for a social network, whose properties we studied through the lens of the computer scientist's glasses.

[9] http://www.flickr.com/

The main problems of voting in social networks are the existence of many weak links, the fact that even the most well-known users are not known by many, and the low voter turnout in general. We are convinced that a *transitive proxy voting* is the best choice in such a setting. The design of a transitive proxy voting system for social networks includes many steps, including the choice of a ballot and a ballot interpretation (e.g.: one vote or many votes, cumulative or approval voting, etc.), and the choice of an appropriate system for counting the votes. The latter includes issues such as delegable vs. undelegable mandates, and the choice of the scoring method.

We propose a system of *transitive proxy voting with exponential damping*, that allows us to frame these questions properly by means of a single parameter. We expect a voting system to be robust to small differences in the choice of the damping factor, or small variations in the users' preferences. We also expect the system to produce an output that depends on the users' choices and not only (or too strongly) on the graph itself. Our experimental results are positive in these aspects.

Future work. We have mentioned but left unexplored a number of paths that can be fruitful to study as future work. First, we did not consider users delegating a varying fraction of their vote; for instance, some users may decide to delegate their whole vote, while others may prefer to delegate only half of their vote and keep the other half for themselves. Second, the choice of limiting the ballot to one vote per person may be appropriate in some circumstances but not in others, and one may want to extend our system to the case of multiple votes, possibly with different weights (as in cumulative voting). Third, the choice of considering only symmetric friendship networks is a restriction that in some cases does not hold (as in Flickr where contacts are directed); although it is easy to re-cast the proposal for this scenario, some of the properties we proved do not hold or need to be reformulated. Fourth, dealing with sybil attacks [17], as well as strategic/untruthful voting in social-network voting systems can by itself be a challenging and interesting research topic. Finally, in a scenario with repeated elections, it may be interesting to understand whether the voting relationship is learnable; under this condition, and assuming a certain accuracy, we may use such an information as an alternative and more precise way to assign the missing votes.

Acknowledgments. We thank Aristides Gionis for helpful discussions and comments on an earlier version of this manuscript. We also thank Adam Rae and an anonymous referee for helpful comments.

9. REFERENCES

[1] K. J. Arrow. *Social Choice and Individual Values*. John Wiley & Sons, 1951, 2nd ed., 1963.

[2] D. Austen-Smith and J. Banks. Monotonicity in electoral systems. *American Political Science Review*, 85(2), June 1991.

[3] P. Boldi, M. Santini, and S. Vigna. PageRank as a function of the damping factor. In *Proc. of the Fourteenth International World Wide Web Conference (WWW 2005)*, pages 557–566, Chiba, Japan, 2005. ACM Press.

[4] E. Elmacioglu and D. Lee. On six degrees of separation in dblp-db and more. *SIGMOD Rec.*, 34(2):33–40, 2005.

[5] B. Ford. Delegative democracy. `http://www.bford.info/deleg/deleg.pdf`, May 2002.

[6] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*, 99(12):7821–7826, June 2002.

[7] J. Golbeck. *Computing with Social Trust*. Springer, 2009.

[8] J. Green-Armytage. An introduction to proxy-based democratic systems. `http://fc.antioch.edu/~james_green-armytage/vm/proxy2007.htm`, Written May 2007, posted online January 2009.

[9] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*.

[10] G. Jeh and J. Widom. Scaling personalized web search. In *In Proceedings of the Twelfth International World Wide Web Conference*, pages 271–279. ACM Press, 2002.

[11] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[12] R. S. Katz. *Democracy and Elections*. Oxford University Press, 1997.

[13] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003)*.

[14] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical Report 66, Stanford University, 1999.

[15] M. A. Rodriguez, D. J. Steinbock, J. H. Watkins, C. Gershenson, J. Bollen, V. Grey, and B. Degraf. Smartocracy: Social networks for collective decision making. *Hawaii International Conference on System Sciences*, 2007.

[16] J. R. Seeley. The net of reciprocal influence: A problem in treating sociometric data. *Canadian Journal of Psychology*, 3:234–240, 1949.

[17] N. Tran, B. Min, J. Li, and L. Submaranian. Sybil-resilient online content voting. In *Proceedings of the 6th Symposium on Networked System Design and Implementation (NSDI)*, 2009.

[18] K. Walsh and E. G. Sirer. Fighting peer-to-peer spam and decoys with object reputation. In *P2PECON '05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 138–143, New York, NY, USA, 2005. ACM Press.

[19] H. Yamakawa, M. Yoshida, and M. Tsuchiya. Toward delegated democracy: Vote by yourself, or trust your network. *International Journal of Human and Social Sciences*, 1(2), 2007.