Non-conservative Diffusion and its Application to Social Network Analysis

Rumi Ghosh, Kristina;
Konstantin Voevodski
Shang-Hua Teng
USC, Boston Univ

Presented by Bing He

Random Walk

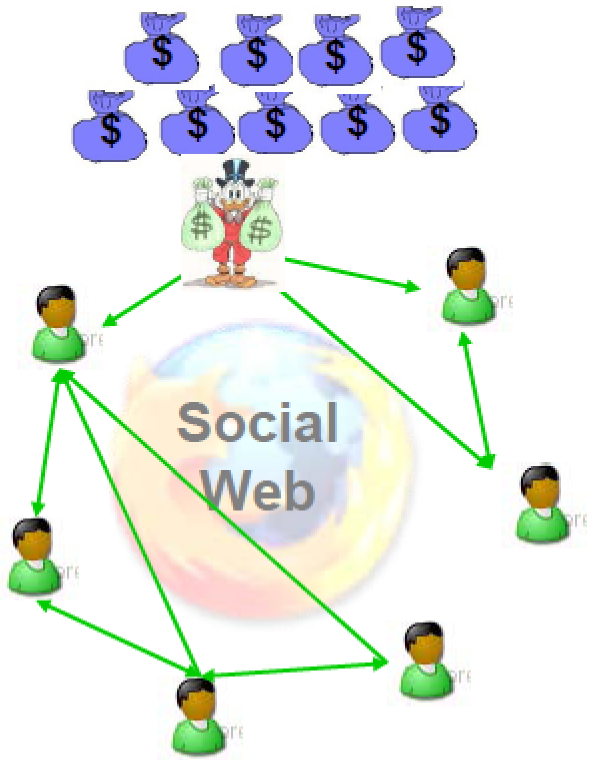
- Simple random walk
- Gaussian random walk
- Correlated random walk
- Random walk on a graph

Conservative and Non-conservative

Network diffusion is a process of the spread of a mass in a network or graph. Mathematically, a diffusion process is given by a function $F : (R^+ \cup \{0\})^{|V|} \rightarrow (R^+ \cup \{0\})^{|V|}$, i.e., a map from a $|V|$-dimensional non-negative vector to a $|V|$-dimensional non-negative vector. This function maps the amount of mass in each node at time t to the amount of mass in each node at time $t+1$. The function F defines a *conservative diffusion* process if for all $w \in (R^+ \cup \{0\})^{|V|}$, $||w||_1 = ||F(w)||_1$, i.e., it preserves the sum of the entries. The function F defines a *non-conservative diffusion* process if for some $w \in (R^+ \cup \{0\})^{|V|}$, $||w||_1 \neq ||F(w)||_1$.

Conservative Process

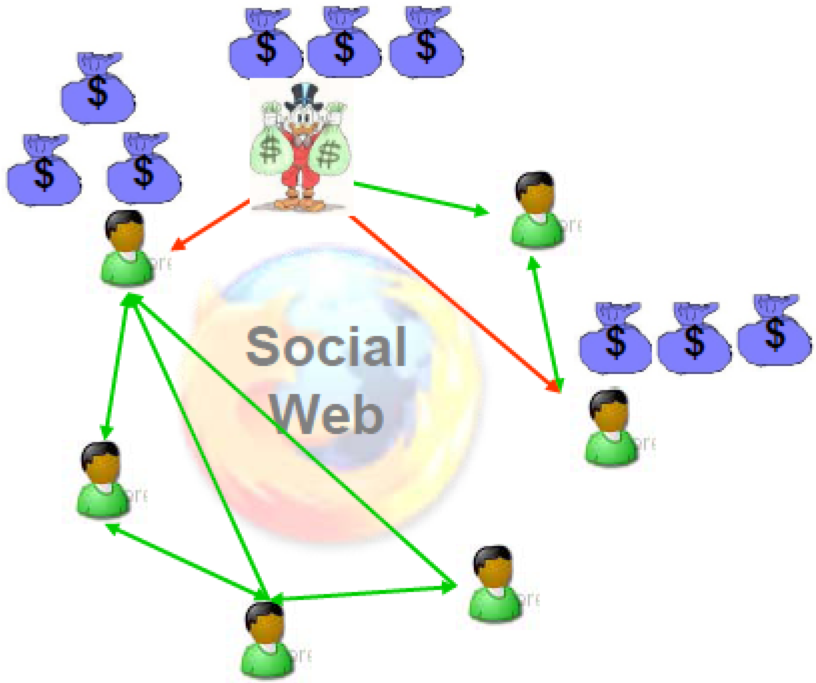
Blue Bags in the network=8



Conservative Process

USC Viterbi School of Engineering

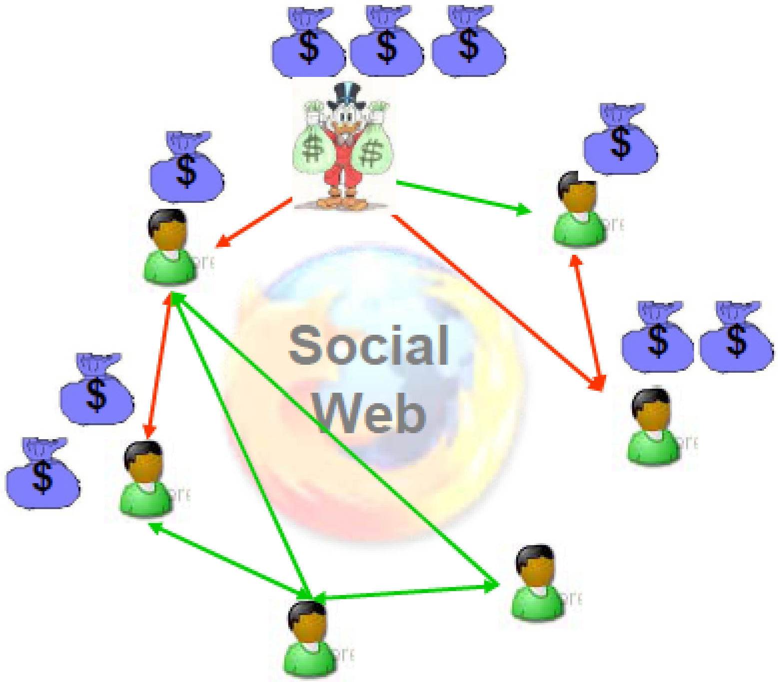
Blue Bags in the network=8



Conservative Process

USC Viterbi School of Engineering

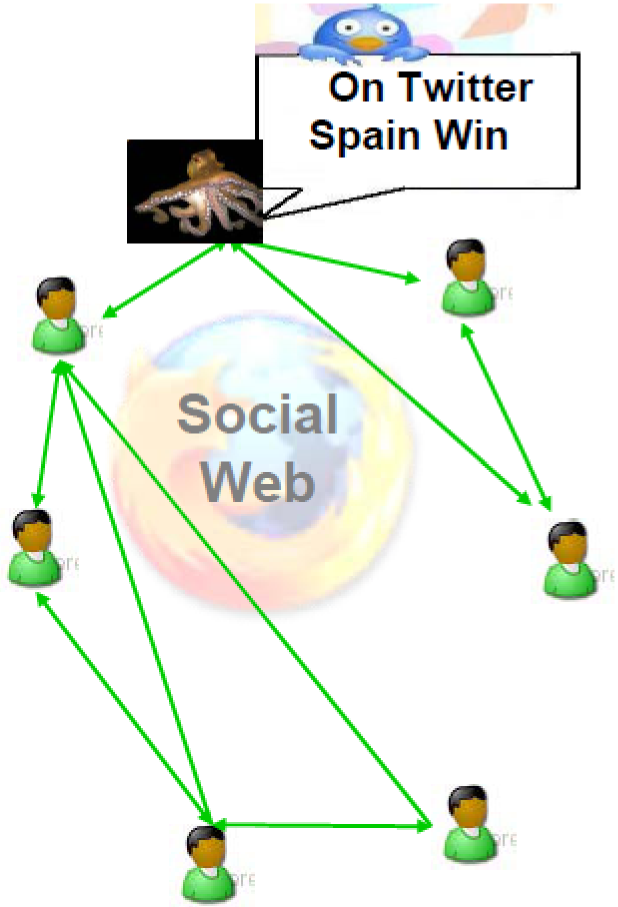
Blue Bags in the network=8



Non-Conservative Process

USC Viterbi School of Engineering

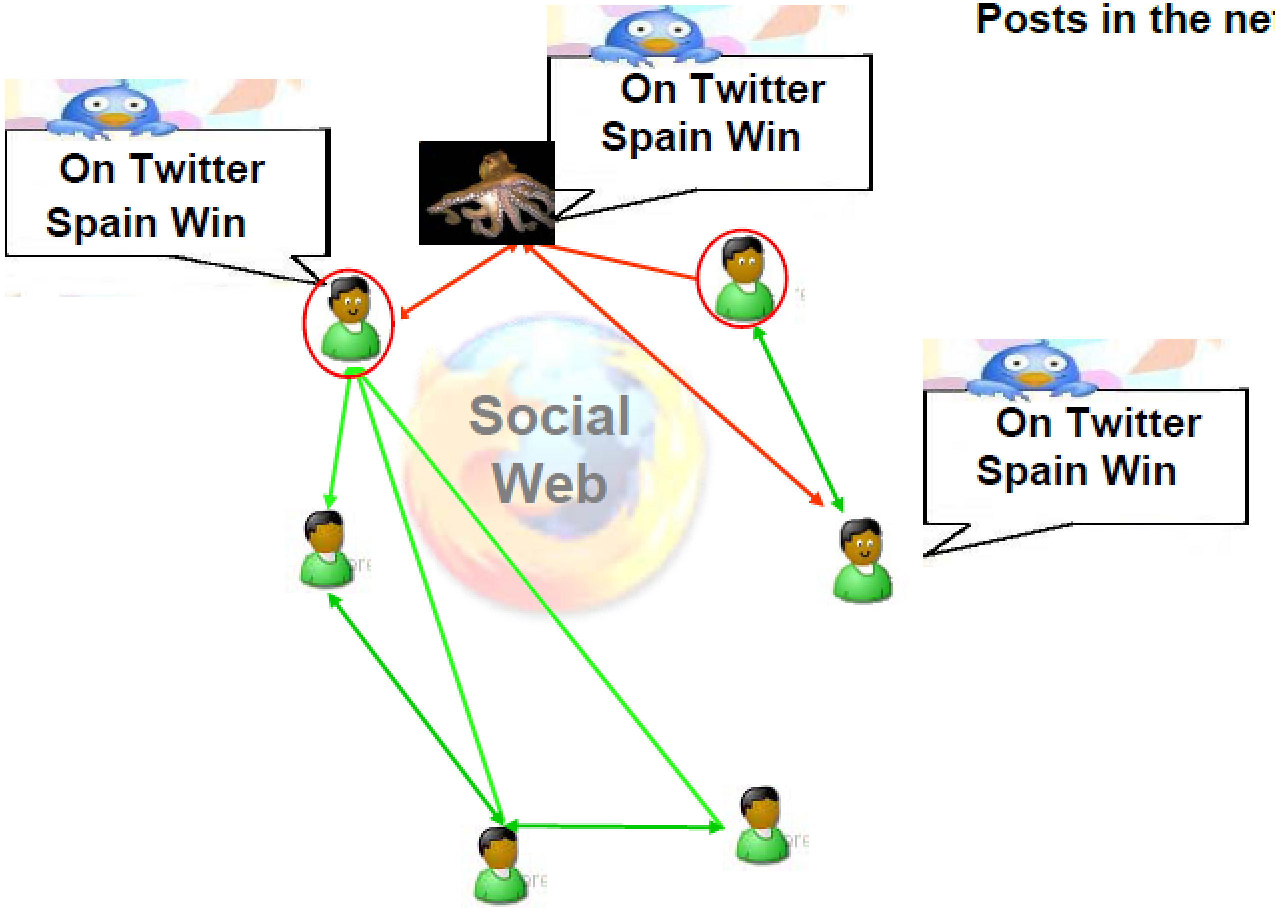
Posts in the network=1



USC

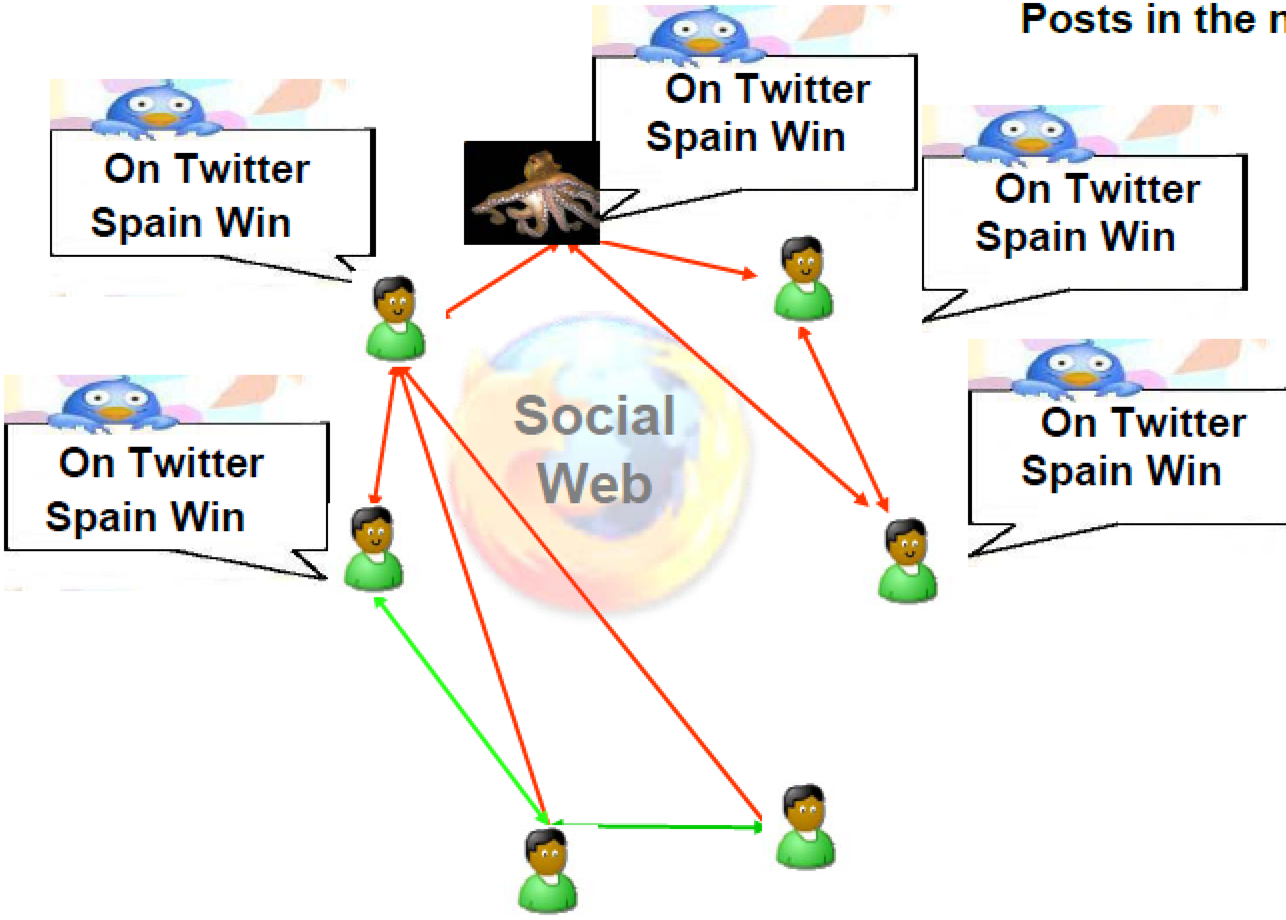
Non-Conservative Process

Posts in the network=3



Non-Conservative Process

Posts in the network=5



Conservative and Non-conservative

- Conservative diffusion simply redistributes the weights among the nodes of the graph, while the total weight remains constant.
  - Examples of conservative diffusion include the flow of money and goods, since the total amount of money and goods remains constant.
- In such non-conservative diffusion, the total weight, or “mass," of nodes in the network is not conserved.
  - Examples of non-conservative diffusion include the spread of information, gossip, innovation, and epidemics.

Alpha-Centrality

- we describe two types of diffusion processes and show that Bonacich's Alpha-Centrality [5] mathematically captures a non-conservative diffusion process.

Alpha-Centrality

$$\mathrm{cr}_\alpha(s) = s + \alpha \cdot \mathrm{cr}_\alpha(s) A. \qquad (2)$$

$$\mathrm{cr}_\alpha(s) = s(I - \alpha \breve{A})^{-1}$$

$$\sum_{t=1}^{\infty} \alpha^t A^t = (I - \alpha \breve{A})^{-1} - I$$

$$\mathrm{cr}_\alpha(s) \;=\; s(I - \alpha A)^{-1} = s \sum_{t=0}^{\infty} \alpha^t A^t. \qquad (3)$$

This formulation shows that the Alpha-Centrality score of node u is given by the number of paths to u from all nodes in the network, exponentially attenuated by their length. The parameter α controls how much weight we give to longer paths. Setting α close to zero enables us to probe only the local structure of the network. As α increases, longer paths become important, and it becomes a more global measure.

Alpha-Centrality

$\frac{1}{\lambda_1}$ is the *radius of convergence* which is the inverse of the spectral radius of the network. A related concept is the *radius of centrality* which is an indicator of how far a message can spread. Taking probability of transmission of a message through an edge to be proportional to α; how far the message will spread or expected path length can be computed by the weighted average of the paths of length of one, two, three and so on, where the weights depend on parameter α. This quantity, which can be shown to be $(1-\alpha)^{-1}$ when $\alpha < \frac{1}{\lambda_1}$, is known as the radius of centrality, and it determines how far, on average, a node's influence will be felt. In Alpha-Centrality, this quantity defines "the radius within which power or centrality is being assessed" [5].

Alpha-Centrality

local or a global metric. For small α, Alpha-Centrality and normalized Alpha-centrality measure local structure of the network and reduce to degree centrality for $\alpha = 0$([6]). As α increases and longer paths become more important, normalized Alpha-Centrality becomes a more global metric, until it hits some fundamental length scale of the network determined by the radius of convergence and given by the expected path length $r_{avg} = (1 - 1/\breve{\lambda}_1)^{-1}$.

Alpha-Centrality vs PageRank

$$\mathrm{pr}_\alpha(s) = (1 - \alpha)s + \alpha \mathrm{pr}_\alpha(s)W.$$

$$\mathrm{cr}_\alpha(s) = s + \alpha \cdot \mathrm{cr}_\alpha(s)A.$$

one:$W = D^{-1}A$. Here the D matrix is the degree matrix, which is a diagonal matrix given by: $D(u, v) = d_{\mathrm{out}}(u)$ if $u = v$; otherwise, $D(u, v) = 0$.

Alpha-Centrality vs PageRank

- The fundamental differences between network diffusion processes impacts the choice of the centrality metric.

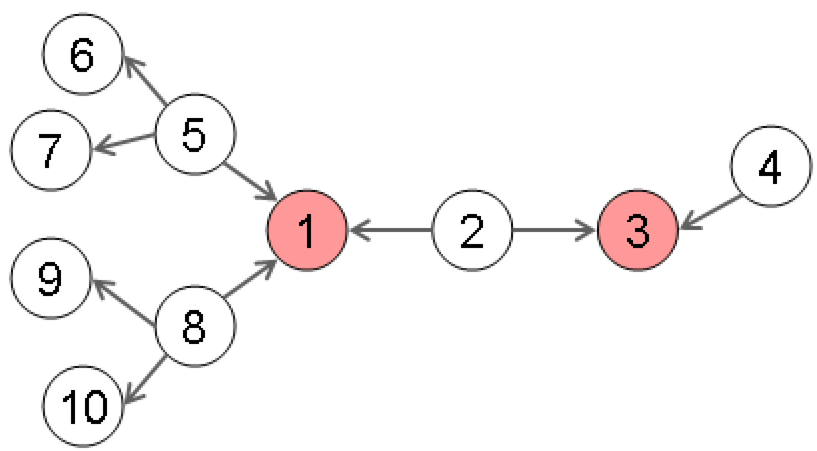


Figure 1: An example network, where node 1 has the highest Alpha-Centrality followed by node 3. In contrast node 3 has the highest PageRank followed by node 1.

Experiment & Evaluation

- Digg
- Twitter
- Standard: we define two empirical measures of influence based on how users react to the information their friends create.

  Estimate 1: Average number of fan votes.

  Estimate 2: Average cascade size.

Results

- Digg
  - We claim that since the spread of news or information is a non-conservative diffusion process, Alpha-Centrality will better reproduce empirically measured influence than a conservative metric (such as PageRank) on Digg, which is a social news aggregator.
- Twitter
  - On the other hand our preliminary investigation of Twitter suggests that its dynamics may be driven by other factors than information spread, for example, conversations [8]; therefore, it may be better modeled by a conservative metric, such as Page Rank.

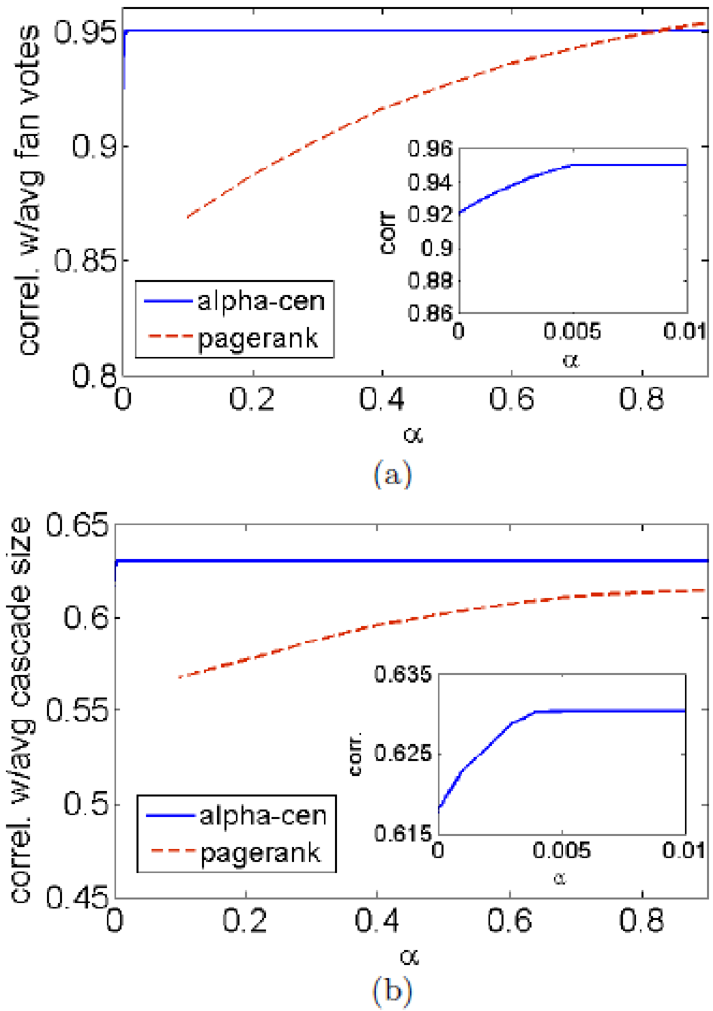


Figure 4: Correlation between the rankings produced by the empirical measures of influence and those predicted by Alpha-Centrality and PageRank. We use (a) the average number of fan votes and (b) average cascade size as the empirical measures of influence. The inset zooms into the variation in correlation for $0 \leq \alpha \leq 0.01$

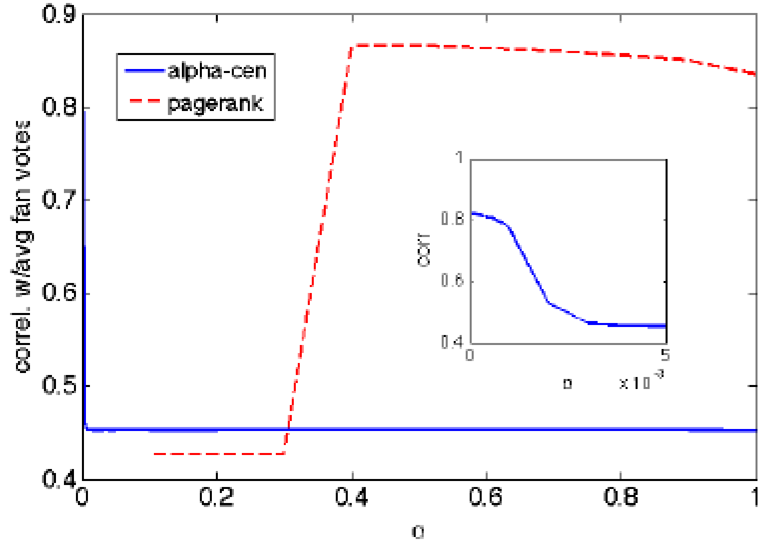


Figure 5: Correlation between the rankings produced by average number of follower retweets and those predicted by Alpha-Centrality and PageRank for Twitter. The inset zooms into the variation in correlation for $0 \leq \alpha \leq 0.005$