Predicting Node Degree Centrality
with the Node Prominence Profile

Yang Yang

Yang Yang, Yuxiao Dong, and Nitesh V. Chawla. Predicting Node Degree Centrality with the Node Prominence Profile. Scientific Reports 4, Article number: 7236, 2014.
Introduction

• In social networks, node and network co-evolve, creating an intertwined effect of centrality and relative position of the node

• Prediction of centrality can lead us to infer importance and/or success of a given individual, which can also provide us important implications of mechanisms underlying social network evolution
Problem Statement

• Problem
  – Here we examine the predictability of centrality of a node in the social network
  – We use the popular degree centrality as a metric for the prediction target (previous studies have found centrality measures to be correlated\(^1\),\(^2\))
  – This task enables us a new perspective to explore the mechanisms driving the evolution of social networks

Background

• Preferential Attachment and Triadic Closure are two principles successfully capturing the main characteristics of social networks
  – Preferential Attachment requires global information
  – Triadic Closure needs local structural information
Methods

• Node Prominence Profile
  – Objective: Reconcile preferential attachment and triadic closure to capture a node’s prominence profile

Node Prominence Profile. In this figure, we mark 5 automorphism positions (labeled with v) in 3 triad structures. The node prominence profile of a node v, is a vector describing the occurrence frequencies of node v in these 5 automorphism positions.
Preliminaries

- **IN and NIN**
  - We partition nodes into important and non-important nodes based on their degree centrality, denoted as **IN** and **NIN**, see Methods and Supporting Information, 1.2)

- **NPP**
  - We call our framework the Node Prominence Profile (NPP)

- **Structural Balance**

\[
\text{balance rate} = \frac{3 \times \text{number of closed triads}}{\text{number of connected triads}}
\]
Results

• Triadic Closure Effect on Degree Centrality Evolution
  – we observe that initially the sub-network of future important (having high degree centrality in future) nodes has a lower balance rate than the sub-network of future non-important nodes, but the former sub-network evolves to form a more balanced topology
Results

• Triadic Closure Effect on Degree Centrality Evolution
  – Implication A: There exist connections between the triadic closure and the degree centrality evolution. In addition, new links are more likely to form between nodes located in an unbalanced sub-network
  – Implication B: The initial sub-network where future important nodes are located is more imbalanced than that of future non-important nodes, thus position of node can be indicative of its future degree centrality
  – These findings are demonstrated to be statistically significant at 95% confidence even if we scale the threshold value of important nodes, such as 10%, 30% and 50% (see Supporting Information, 2.2)
Results

• Positions in Triad Structure
  – We observe that the position of a node within its local structure is related with its degree centrality evolution
  – For instance, triad 2 is more likely to attract new links
Results

• Positions in Triad Structure
  – Different positions of a node in corresponding triads have more descriptive power of degree centrality evolution than existing centrality measures
Results

• Positions in Triad Structure
  – Position 4 is more significantly related with node degree centrality progression
  – nodes in position 4 have high probability to be located in position 3, which means nodes in position 4 are influenced by both mechanisms of preferential attachment and triadic closure
  – This also reflects an important property of these positions in triad structures—they combine the two well known social principles
Results

• Prominence: Centrality and Position
  – In order to demonstrate that prominence is not only represented in the node’s centrality but also in the node’s position in local structure, we provide a detailed investigation into their interaction from the perspective of influence events and provide the evidence that the NPP is able to model both centrality and position information.

Definition  A node \( u \) is said to have a link influence on its neighbor \( v \) iff: 1) there is a link action of node \( u \) with another node \( w \) at time \( t \); 2) there exists a link action of node \( v \) with node \( w \) at time \( t' \); 3) \( \min(T_E(u, v)) < t < t' \) and \( t' - t < \sigma \).
Results

- Prominence: Centrality and Position
Results

• Prominence: Centrality and Position
  – Important nodes have much higher probability to have link influence on their neighbors, and it also validates the principle of preferential attachment (\(|1XX| > |0XX|\) and \(|X1X| > |X0X|\), see figure in last page)
  – Non-important nodes play an important role to transfer link influence (\(|XX0| > |XX1|\))
  – Link influence is more likely to happen between important nodes (\(|11X| > |00X|\))
  – If link influence occurs among important nodes and non-important nodes, then important nodes and non-important nodes have the same chance to initiate the influence (\(|10X| \approx |01X|\))
  – These patterns persist in four different real-world networks and are proved to be statistically significant (see Supporting Information, 2.4)
Prediction of Future Degree Centrality

• Modeled as a classification problem
  – Predict whether new nodes joining the network will be important nodes or non-important nodes in future

Table 1 | Predict Future Degree Centrality. We solve the future degree centrality prediction problem (Supporting Information, 3) using supervised learning method. The five NPP positions (Figure 1) census contributes to our NPP method for prediction. PA (preferential attachment) method just includes the degree centrality feature, and TC (triadic closure) method includes the position 3 as feature. The method labeled All includes existing centrality measures (Supporting Information, 3.1). The supervised learning task is to predict whether a new arriving node will become a important node or a non important node (determined by its degree centrality, see Supporting Information, 3.2) in future. The experiment settings are provided in Supporting Information, 3.2

<table>
<thead>
<tr>
<th>Datasets</th>
<th>AUROC</th>
<th></th>
<th></th>
<th>AUPR</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA</td>
<td>TC</td>
<td>All</td>
<td>NPP</td>
<td>PA</td>
<td>TC</td>
</tr>
<tr>
<td>Condmat</td>
<td>0.85</td>
<td>0.72</td>
<td>0.85</td>
<td><strong>0.86</strong></td>
<td>0.68</td>
<td>0.42</td>
</tr>
<tr>
<td>DBLP</td>
<td>0.79</td>
<td>0.83</td>
<td>0.72</td>
<td><strong>0.85</strong></td>
<td>0.27</td>
<td>0.34</td>
</tr>
<tr>
<td>Enron</td>
<td>0.71</td>
<td>0.55</td>
<td>0.70</td>
<td><strong>0.72</strong></td>
<td>0.43</td>
<td>0.18</td>
</tr>
<tr>
<td>Facebook</td>
<td><strong>0.81</strong></td>
<td>0.78</td>
<td>0.74</td>
<td><strong>0.81</strong></td>
<td>0.42</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Prediction of Future Degree Centrality

- Modeled as a regression task
Prediction of Future Degree Centrality

• Generality of Node Prominence Profile: Prediction Across Datasets
  – This indicates that the prominence profile of node captures principles that are more generic than the state-of-the-art centrality measures, and this still holds even if the generalization is across different domains of networks

![Graph a](image1.png)
![Graph b](image2.png)
Conclusions

• We demonstrated that the position of a node in a local structure is strongly indicative of the degree centrality progression in the social network
• Building on this observation, we developed a prediction method referred to as NPP. We empirically demonstrated the effectiveness of NPP by demonstrating improvement in performance over the state-of-art methods
• The generalization capacity of our methods under a transfer learning scenario showed that our approach is able to capture essential properties or features underlying degree centrality evolution
• Our findings have important implications for the applications that require prediction of a node’s future degree centrality, as well as the study of social network dynamics