

Centrality Fairness: Measuring and Analyzing Structural Inequality of Online Social Network

Heeyoung Kwak¹, Joonyoung Kim¹, Yongsub Lim², Shin-Kap Han³, Kyomin Jung¹
¹Department of Electrical and Computer Engineering, Seoul National University, Korea
²Department of Computer Science and Engineering, Seoul National University, Korea
³Department of Sociology, Seoul National University, Korea
 {hykwak88, kimjymcl, yongsub, shinkaphan, kjung}@snu.ac.kr

Abstract

While measuring inequality of a social system has been a popular topic in economics and sociology, structural fairness and inequality of social networks has not been paid attention by researchers interested in web or social network analysis. In practice, measuring structural fairness and inequality has a number of applications in online social networks, for example, we can check skewness of degree distribution by simply seeing inequality index. The power-law exponent has often been used to measure the inequality of network structures, however, it has several drawbacks to be applied to universal networks.

In this paper, we propose a novel framework to measure fairness and inequality of a given network in the context of its structure. We develop a set of centrality fairness measures by combining other well-known node centralities with Gini index. We also analyze scale-free property of our proposed centrality fairness measures in real networks.

Moreover, we suggest simple and efficient methods to relax structural inequality of a network, which are based on two edge manipulations: addition and rotation. Through experiments on real networks, we show that our methods decrease inequality quite steadily and effectively, and as structural hierarchy of a network gets stronger, decreasing rate of inequality gets lower.

Keywords: Centrality fairness, Structural inequality, Node centrality, Gini index, Social network.

1 Introduction

In economics and sociology, inequality of a real world system has been studied with respect to various dimensions such as income [1], education [2], and opportunity [3]. With the recent growth of web and mobile contents such as blogs and social network services, online social networks have been spotlighted as one of rich information sources reflecting the real world social system. However, despite a number of studies on measuring a network characteristic such as clustering coefficient and modularity, it has been hardly done in terms of structural inequality even though

“structural” and “inequality” easily come up with word “network” and “ocial,” respectively.

In traditional inequality researches, a system is viewed as collection of individuals without considering relation or interaction between them. For such a reason, it has not been particularly relevant to measuring how structurally unequal the system is. Thus, existing measures cannot be directly applied to measuring structural inequality of a network.

In a real world network system, measuring fairness and inequality in terms of its structure can provide an idea in characterizing a network or controlling its growth to keep the network somewhat structurally equal. For example, in a social network service like Twitter, considering inequality of indegree and/or outdegree may enable us to know that vitalization of the system highly depends on a small number of power users, and try to develop contents actively consumed within passive users. Another example is a peer-to-peer (P2P) network. In this case, a developer or manager would like the network to grow or shrink while keeping somewhat high equality in betweenness or closeness for efficient transmission. In fact, even without measuring inequality, one can handle the cases above in some way. However, by summarizing the extent of inequality, we can characterize a structural state of a network by some simple indices, which can be of use to prevent/help that the network changes to an undesirable/desirable structure.

In general, the power-law exponent has been used to measure the fairness of social networks and complex network systems. However, due to several weaknesses of this approach, the power-law exponent has significant limitations as a fairness measure, particularly narrow applicability [4]. Many real-world networks do not follow a power law and it is not valid to use power-law analysis for these networks. For example, some network structures like ad hoc network are best captured by a random geometric graphs which have Poisson degree distributions, thus do not exhibit scale-free properties. Also, the power-law exponent suffers from high computational complexity and the power-law paradox i.e., the power-law exponent does not precisely quantify the extent of inequality/unfairness of the degree distribution. As alternatives, there have been proposed several measures based on the Lorenz curve, which is

widely used in economics. The recent paper shows that the Gini coefficient of a degree distribution does not suffer under the drawbacks of the power-law exponent [4].

In this paper, we first propose a framework for measuring fairness of a network of other structural components of a network besides degree distribution. In our framework, structural wealth of individuals is quantified by a node centrality such as pagerank, closeness and betweenness and then the wealth distribution is summarized as one value by Gini index. Combining known node centralities and Gini index, we develop a set of measures called centrality fairness, each of which we examine its meaning in terms of information spreading. Also our measures can be considered as a kind of network characteristic measures such as the exponent of a power-law degree distribution or clustering coefficient.

Next, we analyze the scale-free property of our proposed structural inequality measures in real networks. A network is said to have a scale-free property when the characteristic of whole network also holds in a part of the network. A scale-free property is essential for the inequality measure in that we can infer the whole network structure by only seeing a small part. Thus, we measure correlation of inequalities between the whole network and sampled subset. Especially we focus on nodes having high centralities (rich group) as the subset. This rich group does not show exactly scale-free property, but we observe that centrality fairness is positively correlated with that of the whole network.

Lastly, we suggest methods to relax structural inequality of a network and verify the results using centrality fairness. Especially we focus on a process manipulating relation of individuals, i.e., edges, and apply two types of manipulations: edge addition (making new relation) and edge rotation (changing existing relation). Through experiments, both methods reduce inequality steadily on average over the amount of manipulations, and the rotation shows more stable performance. We also observe that decreasing rate of inequality shows negative correlation with strength of structural hierarchy of a network. Thus, we analyze the structural characteristics that enable each method to greatly reduce inequality even with quite small manipulations.

1.1 Our Contribution

We summarize our contributions as follows.

- We devise a new framework to measure structural fairness and inequality of a network. Especially, our proposed measures can cover the non-power law networks. Our method essentially combines node centralities and Gini index, each of which has its own meaning depending on the used centrality.

- We analyze scale-free property of our proposed structural inequality measures in real networks. By measuring correlation of inequalities between the whole network and sampled subset, we verify the scale-free property in our proposed measures.
- We suggest network manipulation methods for relaxing structural inequality based on two elementary operations: edge addition and rotation. Through experiments, we show that both methods reduce inequality steadily on average over the amount of manipulations, and decreasing rate of inequality is negatively correlated with strength of structural hierarchy of a network.

1.2 Related Work

There are a large number of studies, especially in economics and sociology, measuring inequality between individuals in various dimensions, such as income. The key issue in these studies is how to summarize a given distribution of individuals' wealth by a single measure to indicate the extent of inequality [5-6].

The framework popularly used is Lorenz curve, which is a kind of normalized cumulative distributions and becomes the $y = x$ line for a perfectly equal distribution. Of measures based on Lorenz curve, the most commonly used one is Gini index [7]. This is defined as the area between the perfectly equal line and Lorenz curve, which becomes large when a small group possesses a disproportionately large portion of the total.

Based on Lorenz curve, some network fairness measures are proposed in [4] to analyze the structure of the Web, resolving the weaknesses of power-law exponent. However, this work only focuses on measuring unfairness of the degree distribution. The paper quantifies the fairness of degree distribution, one of the known node centralities, using Gini index but it lacks measures for the fairness of other node centralities such as pagerank, closeness and betweenness.

Within the social network literature, the "centralization" proposed by [8] is one of structural inequality measures. Given a centrality such as betweenness, it measures difference between the most central node of a given network and that of the worst case network (i.e., with the theoretically largest such sum of differences) with the same number of nodes [9]. For our purpose here, these works lack consideration of the whole distribution, and more importantly, finding the worst case network is problematic in working with various node importance measures.

A recent study also analyzed structural features of networks to propose a learning model for mining influential communities [10]. In another recent study, structural analysis is applied to describe a certain real-world network. [11] characterizes the structural features of the Twitter

follow graph in terms of graph properties such as degree distributions, clustering coefficients, and shortest-path lengths.

2 Setting and Preliminaries

In this section, we describe the network datasets used in our analysis, and explain a node centrality and Gini index.

2.1 Data Description

We use nine directed networks, including online social networks, blogs, web links, trust networks, and voting. All the networks below are manipulated so that isolated nodes and duplicated edges are removed.

- advogato [12]: A trust network of Advogato, SNS for open source developers. Trust is represented by one of values $\{0.4, 0.6, 0.8, 1.0\}$, and we excluded edges with 0.4 which stands for a less-than-sufficient level in the original paper.
- anybeat [13]: A follower/followee network(as Twitter) of a social network named Anybeat. The following relationships are represented by directed edges.
- google [14]: A hyperlink network of Googles own webpages. International pages and nodes farther than 3 steps from the start node are excluded.
- polblogs [15]: A hyperlink network of weblogs among political-oriented blogs during the 2004 U.S. Presidential campaign. A link is constructed if a URL present on the page of one blog references another political blog.
- slashdot-re [16]: A reply network of Slashdot which is a technology-related news website. An edge is from a replier to an author of a post.
- epinions [17]: A trust network of Epinions in which edges are labeled by trust or distrust. We used only trust edges whose proportion over the total is 0.85.

- slashdot [17]: A trust network of Slashdot in which edges are labeled by friend or foe. We used only friend edges whose proportion over the total is 0.77.
- wiki-vote [18]: A wikipedia voting network for promoting to an administrator in which edges are labeled by positive or negative. We used only positive edges whose proportion over the total is 0.79.
- facebook [19]: A collection of wall posts from the Facebook New Orlean networks. A posting from a users friends is treated as a form of user interaction.

2.2 Node Centrality

Node centralities have been studied in social network analysis to rank nodes with respect to a certain importance criterion. For instance, the indegree and outdegree centrality measures how popular and active each node is in a network, respectively; the pagerank measures how many influential nodes a node is linked from; the betweenness centrality measures how often a node is used in connecting other node pairs with a shortest path; the closeness centrality measures how fast, in terms of the number of hops, a node can reach all nodes in a network; Below, we present formal definitions of these centralities. In this paper, we use the notations $G = (V, E)$ with $n = |V|$ and $m = |E|$ for a network, and $d_{in}(u)$ and $d_{out}(u)$ for indegree and outdegree of a node $u \in V$, respectively.

- Degree: size of indegree (although we focus on indegree, it can be applied to outdegree).

$$C_{deg}(u) = d_{in}(u). \quad (1)$$

- Betweenness: the extent that each node is placed on shortest paths between other nodes.

$$C_{bnw}(u) = \sum_{v, w \in V \setminus \{u\}} \frac{|\sigma_{vw}(u)|}{|\sigma_{vw}|}, \quad (2)$$

Table 1 Statistics of Our Datasets

Name	$ V $	$ E $	λ	α_{in}	α_{out}	α_{all}	ρ	Δ	κ	ω
advogato	5,170	47,334	91.16	1.46	1.49	1.57	0.09	9.00	0.98	0.00
anybeat	12,645	67,053	5.30	1.61	1.70	1.90	0.02	10.00	1.00	0.00
google	15,763	170,335	10.81	1.48	1.38	1.48	0.01	7.00	1.00	0.00
polblogs	1,224	19,022	15.54	1.42	1.34	1.40	0.23	8.00	1.00	0.00
slashdot-re	51,083	130,370	2.55	1.99	1.70	2.34	0.01	17.00	1.00	0.07
epinions	114,222	717,129	6.28	1.76	1.68	2.04	0.09	14.00	0.88	0.00
slashdot	75,144	425,072	5.66	1.62	1.72	1.85	0.03	14.00	0.99	0.00
wiki-vote	6,262	81,820	13.07	1.66	1.47	1.58	0.13	8.00	0.99	0.00
facebook	45,813	264,004	5.76	1.50	1.52	1.64	0.09	18.00	0.96	2.24

Note. In the first row, λ stands for density i.e., $|E| = |V|$; α_{in} , α_{out} and α_{all} for exponents of the power law curve for indegree, outdegree, and both, respectively; ρ for clustering coefficient; Δ for diameter; κ for size of maximal connected component; ω for ratio of multiple edges.

where σ_{vw} is a set of shortest paths from v to w , and $\sigma_{vw}(u) = \{(s_1, \dots, s_k) \in \sigma_{vw} : \exists i, s_i = u\}$.

- Closeness: the average shortest path length from a node to the other nodes.

$$C_{cls}(u) = \sum_{v \in V} 2^{-d(u,v)}, \tag{3}$$

where $d(u, v)$ is the shortest path length from u to v (we use the definition from [20] for a weakly connected network).

- Pagerank: higher scores as pageranks of in-neighbors gets higher.

$$C_{pgr}(u) = 0.15 + 0.85 \sum_{(v,w) \in R} \frac{C_{pgr}(v)}{d_{out}(v)}, \tag{4}$$

where we use 0.85 as the damping factor.

2.3 Gini Index

Gini index is one of the most popular inequality measures which has been applied in various fields, including sociology [21], economics [3], and ecology [22]. The definition of Gini index is as follows.

Definition 1 (Gini index). Given a vector $X \in \mathbb{R}^n$, let Y be a sorted vector of X in increasing order. Then, the Lorenz curve $L: [0, 1] \rightarrow [0, 1]$ is defined as a piecewise linear function connecting $(x(k), l(k))$, $0 \leq k \leq n$ where

$$x(k) = \frac{k}{n}, l(k) = \frac{\sum_{i=1}^k Y_i}{\sum_{i=1}^n Y_i}. \tag{5}$$

Then Gini index is defined as

$$\Gamma(X) = 1 - 2 \int_0^1 L(x) dx. \tag{6}$$

Figure 1(a) depicts relation of Lorenz curve and Gini index. Note that for the perfectly equal case, i.e., all elements in X are identical, its Lorenz curve is the $y = x$ line, and for the case of only one nonzero element in X with $n \rightarrow \infty$, the curve goes to the $y = 0$ line. Thus, note that a Lorenz curve is always below the perfectly equal line.

One might be confused between fairness and small variance of wealth. We notice that a small Gini index does not mean that wealth of people has small variance (concentrated to the mean). Rather, it means that a large portion of the total wealth is shared by a majority of the total population. For example, we can consider two societies having wealth distributions of the same variance but very different Gini index, as follows. In one society, a majority is rich and the few is extremely poor; in the other society, a majority is poor and the few is extremely rich.

Corresponding Lorenz curves for both societies are depicted in Figure 1(b). Note that despite the same variance, there is a very large gap in terms of inequality between two societies.

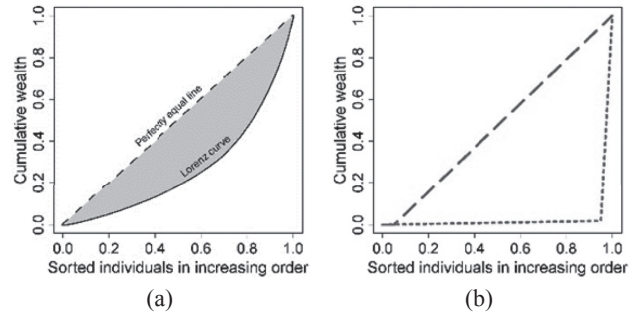


Figure 1 Examples of Lorenz Curve

Note. (a) The shaded area is the half of Gini index. (b) The red dash curve -- extremely equal -- corresponds to the society where only few people is very poor, and the blue dot curve -- extremely unequal -- corresponds to the society where only few people is very rich. Both societies have the same variance, but completely different Gini index.

3 Centrality Fairness Measure

In this section, we propose a framework to measure structural fairness and inequality of a network. Our framework combines a node centrality and Gini index, that is, it uses a centrality as individuals' structural wealth and Gini index to summarize it. One advantage of using Gini index is that it satisfies several properties required to be a good fairness measure such as mean independence, population independence, transfer principle and symmetry [23]. Also it has no parameter, which provides simplicity beneficial when working with a new topic and helps to concentrate on the problem itself.

The exponent of power-law degree (or other centrality) distribution, familiar in the literature, might be considered for a summarization method. In this case, as the value gets larger, a network gets more equal. However, it can be only used when a wealth distribution follows power-law, preventing to apply with general node centrality measures. Also it cannot distinguish the cases of the same exponent but different size of a support set as in Figure 2. Obviously the second case is more unequal than the first one.

Combining the node centralities in the previous section with Gini index, we develop four different centrality fairness measures for a network, and we will denote each measure by concatenating a node centrality name and Gini, e.g., pagerank-Gini. Then, those enable us to examine structural inequality of a network from various angles. Below, we present meanings of our measures in terms of information spreading/transmission.

- Degree-Gini: Degree centrality can be considered

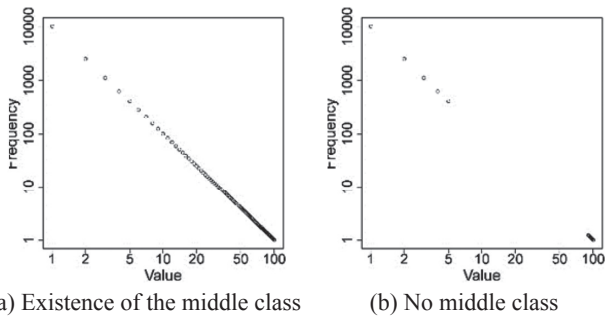


Figure 2 The Same Power Law Distribution with Different Support Sets

to measure how much information a node directly receives, that is, from its direct neighbors. If we consider information attenuated per transmission such as rumor, degree-Gini represents inequality on the ability to receive credible information.

- **Betweenness-Gini:** In information transmission view, betweenness of a node can be interpreted as likelihood of the node being a bottleneck. Then, betweenness-Gini can be an indicator for how many and strong bottlenecks are in a network. In this case, a large value implies existence of few but strong bottlenecks in a network. On the other hand, large betweenness-Gini also means that a large number of shortcuts are spanned by a small number of nodes. Hence, betweenness-Gini can be used as an indicator for how robust a network is to node removal attacks in terms of efficient transmission.
- **Closeness-Gini:** Closeness measures the ability of a node to spread information within a small number of hops to the whole network when the information is originated at that node. Then, large closeness-Gini indicates that there are few effective nodes and many ineffective nodes as the origin of spreading information.
- **Pagerank-Gini:** Considering that pagerank is defined as the stationary distribution of a transition matrix, pagerank measures the amount of information reaching the node through the corresponding random walk. Hence, pagerank-Gini can be interpreted as inequality on the ability to receive information.

Through experiments on our datasets, we observed that in general, betweenness-Gini is very high (at least 0.8), and closeness-Gini was quite low (at most 0.5). The former is because a range of the betweenness centrality is from 0 to n^2 , which can make a rich-poor gap large, and also because of existence of nodes having very large betweenness: hubs and bridges. The latter is from a property of the closeness centrality. First let us consider an undirected network, and assume that a node with the highest closeness can reach all the other nodes within $O(\log n)$ hops. Then, it means that every node can reach all the other nodes within $O(\log n)$ hops. In other words, a node with high closeness itself has a

positive effect to other nodes in reachability. This argument can be applied to directed networks, and thus closeness-Gini appears somewhat low.

We also examined correlations between our fairness measures. As expected, degree-Gini, pagerank-Gini, and betweenness-Gini are somewhat positively correlated each other (about 0.5 ~ 0.8). In contrast, closeness-Gini is weakly correlated with degree-Gini and pagerank-Gini (less than 0.2). Moreover, it shows negative correlation with betweenness-Gini (about -0.55). These moderate correlations, especially between closeness-Gini and the others, indicate that each measure captures structural inequality of a network in a distinct dimension.

Test of Scale-Free Property

In general, a network is said to have a scale-free property when the characteristic of whole network also holds in a part of the network. We analyze scale-free property of our proposed centrality measures in real networks by examining whether inequality of the whole network is also observed in its subset consisting of rich nodes in the centrality. (scale-free property) To that end, we construct a reduced network consisting of a proportion of richest nodes in the centrality, and compare its inequality with that of the whole network.

In Figure 3, correlations of our inequality measures between the whole networks and reduced networks consisting of top 10% nodes in specified centralities are shown. In the figure, correlation forming the $y = x$ line implies that the scale free property holds.

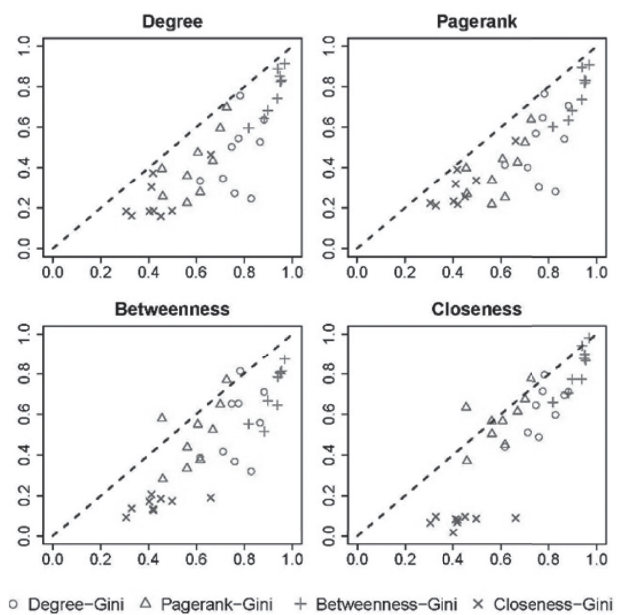


Figure 3 Correlations of Inequality between the Original Networks and Reduced Networks

Note. These graphs show correlations of inequality between the original networks and reduced networks by top 10% nodes in the specified centralities. x-axis is the original inequality and y-axis is inequality in the reduced networks. The nine points for each color are our network datasets.

We observed for all cases that correlations are quite large for degree-Gini, pagerank-Gini and betweenness-Gini, but not that clear for closeness-Gini. Another observation is that inequalities of the reduced networks are larger than that of the whole networks -- correlation slope larger than 1. This makes sense because rich nodes generally have large in- or outdegree and thus such a reduced network becomes denser than the original.

4 Methods to Relax Structural Inequality

We have proposed structural fairness and inequality measures for a network. In many real problems, we would like to design and manage a network to have more fair and equal structure. In this section, we propose methods to relax structural inequality of a given network and analyze them. Our methods are random processes that gradually manipulate edges of a network to make it have a smaller pagerank-Gini. Although we especially focus on pagerank-Gini in this section, our analysis can be naturally applied to our other centrality fairness measures.

Our methods use two types of network manipulations: edge addition and edge rotation. Even though the rotation may be less applicable to real world systems than the addition, it can be of help in designing a network, managing its growth, or understanding a basis of relaxing inequality.

4.1 Edge Addition

Naturally, we can expect that a network gets more equal in pagerank-Gini as more edges are added because a pagerank distribution tends to be uniform as a network gets denser. For addition, we consider a random process determined by two probability distributions over nodes for choosing head and tail nodes of a new edge. Below, we present several methods.

The first method is to choose tail and head nodes uniformly at random, and connect them. This is used as a baseline to see how effectively other additions perform. We call this method rAdd. Considering pagerank of one node distributed over its neighbors along links, we can design a more efficient method. The key idea is to add an edge from a rich node to a poor one. We suggest two methods for each of choosing a tail node (rich) and head node (poor).

For a tail node, the first one is to simply pick a node with probability proportional to its pagerank. The second is to pick a node with probability proportional to the pagerank divided by its outdegree, which focuses on influence of a node spread to its each out-neighbor rather than an absolute pagerank value. We denote these two methods by RichTail and InfluTail, respectively.

For a head node, the first method is to pick a node

with probability proportional to inverse of its pagerank, and the second one is to pick a node with probability proportional to inverse of summation of pagerank of the node and discounted pageranks of its out-neighbors, i.e.,

$$\frac{1}{C_{pgr}(u) + 0.85 \sum_{v \in N_{out}(u)} C_{pgr}(v) / d_{out}(u)}.$$

We denote these by PoorHead and PoorArea, respectively.

4.2 Effect of Edge Addition

Figure 4(a) shows changes of pagerank-Gini as the number of added edges gets larger. As expected, pagerank-Gini decreases as edges are added more, but for some networks, especially slashdot and wiki-vote, it increases with rAdd until $|V|$ number of edges are added. One may think it depends on initial pagerank-Gini or density. But, in our experiments, facebook and advogato, having almost same initial pagerank-Gini to slashdot and wiki-vote, do not show a similar pattern. Rather, we guessed that it is more likely due to a hierarchical characteristic of a network. We use the definition by [24] for a hierarchical structure h^* where as $h^*(u)$ gets larger, $u \in V$ is considered to be on a higher level in the hierarchy. Details on h^* are provided in Appendix. Defining back-edges as $\{(u, v) \in E: h^*(u) - h^*(v) > 0\}$, note that as back-edges gets smaller, structural hierarchy of a network gets stronger. In Figure 5(a), correlations of pagerank-Gini after rAdd with density and ratio of back-edges are shown. Note that decrease ratio of pagerank-Gini is highly correlated with ratio of back-edges, i.e., negatively with strength of hierarchy, and its correlation is clearer than that with the density of a network.

Another notable point in Figure 4(a) is that for google, RichTail \oplus PoorArea and InfluTail \oplus PoorArea rapidly decrease the pagerank-Gini even with a small number of additions compared with the other additions. One reason is that in google, a large portion of poor nodes in pagerank have out-neighbors such that the sum of their pageranks is large as in Figure 6. This means that when we choose a head node for addition using inverse of an absolute pagerank, flow along the new edge spread to a small number of rich nodes or a large number of poor nodes. The former case does not help poor nodes get richer, and the latter case helps it quite insignificantly.

Overall, RichTail \oplus PoorArea and InfluTail \oplus PoorArea are better for a small number of manipulations, but as the amount of manipulations gets larger, RichTail \oplus PoorHead and InfluTail \oplus PoorHead decrease inequality better in general.

4.3 Edge Rotation

We define the rotation of an edge as changing its head node with keeping the same tail node. In contrast to

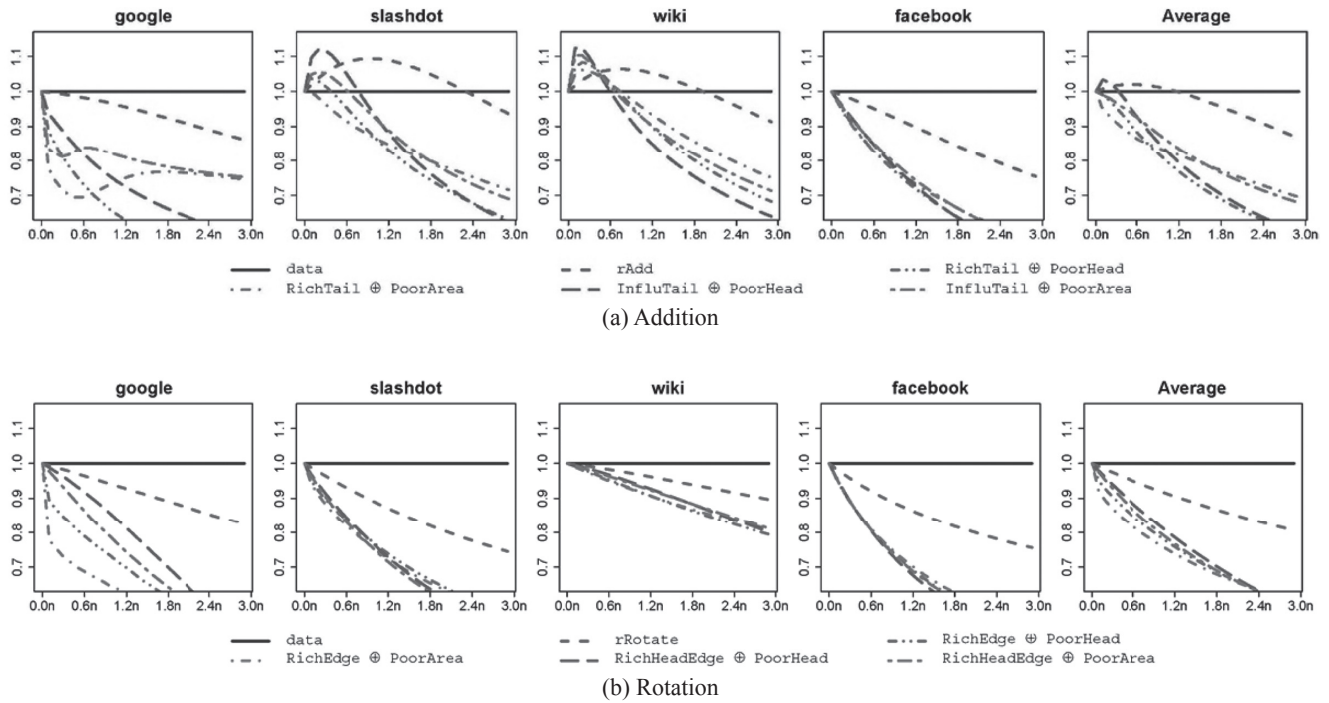


Figure 4 Changes in Normalized Pagerank-Gini in Relation to Changes in Edges

Note. Graphs showing how the addition and rotation methods relax the pagerank-Gini of our selected networks. For each graph, x-axis is the number of added or rotated edges and y-axis is normalized pagerank-Gini by the initial value (data) where $n = |V|$. Although only selected networks are presented here due to the space limit, the other networks showed similar patterns to the average of each case.

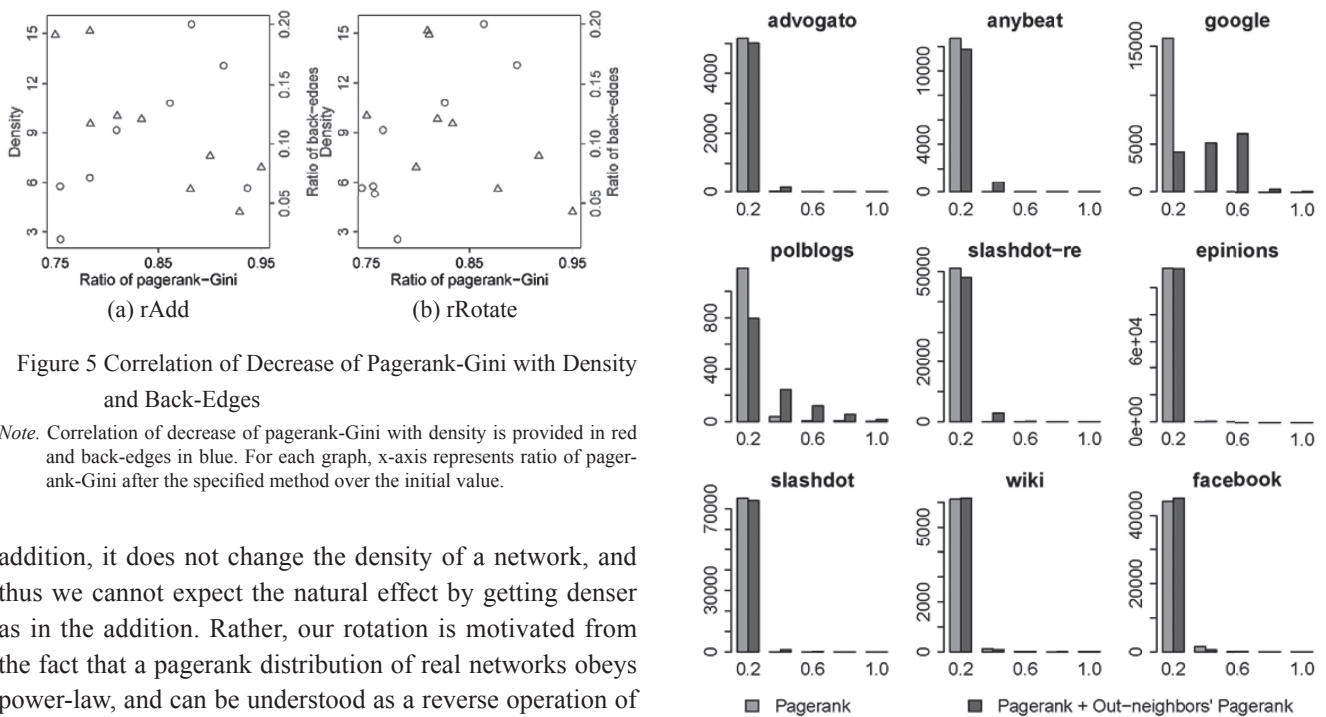


Figure 5 Correlation of Decrease of Pagerank-Gini with Density and Back-Edges

Note. Correlation of decrease of pagerank-Gini with density is provided in red and back-edges in blue. For each graph, x-axis represents ratio of pagerank-Gini after the specified method over the initial value.

addition, it does not change the density of a network, and thus we cannot expect the natural effect by getting denser as in the addition. Rather, our rotation is motivated from the fact that a pagerank distribution of real networks obeys power-law, and can be understood as a reverse operation of the preferential attachment.

We propose one base line and two methods for each of choosing a rotated edge and a new head node. First, as a baseline, we consider to choose a rotated edge and a new head node uniformly at random. Second, for a rotated edge, one is to pick an edge with probability proportional to sum of pageranks of its two end nodes, and the other is to pick it with probability proportional to pagerank of its

head node. We denote these two methods by RichEdge and RichHeadEdge, respectively. The methods for choosing a new head node are the same as those used in the addition.

Figure 6 Histograms Showing Distributions of Pagerank Values and Summation of Pageranks of Out-Neighbors

4.4 Effect of Edge Rotation

Figure 4(b) shows changes of pagerank-Gini as the number of rotated edges gets larger. In contrast to the addition cases, every method monotonically and constantly decreases the pagerank-Gini. Note that for google the value significantly decreases by RichEdge \oplus PoorArea as in the addition cases, but not with RichHeadEdge \oplus PoorArea. This is due to the fact that in google, there are many edges from very poor nodes to very rich nodes compared with other networks. Figure 7 shows that for many edges (u, v) of google having high probability in RichHeadEdge, i.e., having large $C_{pgr}(v)$, $C_{pgr}(v) - C_{pgr}(u)$ is large compared with other networks. This can be regarded as a characteristic of internal networks with centralized management compared with those growing in fully decentralized mechanism.

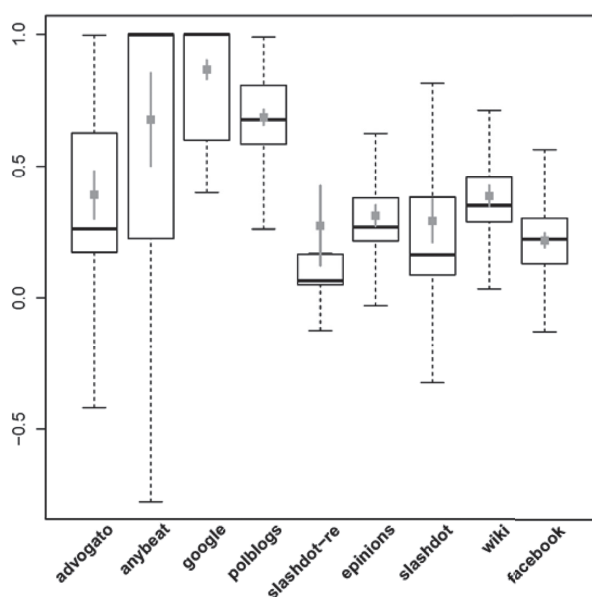


Figure 7 Differences of Pageranks between Head and Tail Nodes of Edges Having High Probability in RichHeadEdge

Note. For each network, top 10% edges are used. The black box plot represents minimum, lower quartile, median, upper quartile, and maximum while excluding outliers outside interquartile range times 1.5 from the box; red bar plot represents mean and variance.

Figure 5(b) shows that performance of rRotate is somewhat correlated with not only the number of backedges (described in Effect of Edge Addition) as in rAdd but also the density of a network. However, note that rRotate never increases pagerank-Gini even with few rotations, and in fact, this is the same for all the other rotation methods. As a consequence, a rotation is more robust to a network structure when relaxing inequality of pagerank-Gini.

5 Conclusion

This paper has proposed a framework for measuring

structural fairness and inequality of a network with a set of measures, analyzed the property of our measures, and developed methods to relax the inequality of a given network. We examined the meanings of our measures in terms of information transmission and the correlations between our fairness measures. Also we observed that fairness of the whole network is similar to the fairness of its subset consisting of rich nodes in the centrality. Our inequality relaxation methods were shown to relax the inequality quite steadily and effectively. We also showed that difficulty of relaxing the inequality is positively and highly correlated with hierarchical strength of a network.

We believe that this paper presents novel measures for network analysis, and can be considered as a starting point for further research on measuring structural fairness and inequality of a network. Especially, we expect that methods to relax (or possibly intensify) inequality has many applications in designing and managing a network and would be one of topics receiving attention in inequality studies of a network. Also, reconstruction of inequality in real world by developing a network generation model with our measures might be addressed in future research.

References

- [1] C. Wang, K. Caminada and K. Goudswaard, The Redistributive Effect of Social Transfer Programmes and Taxes: A Decomposition Across Countries, *International Social Security Review*, Vol. 65, No. 3, pp. 27-48, July-September, 2012.
- [2] V. Thomas, Y. Wang, X. Fan and World Bank Institute, *Measuring Education Inequality: Gini Coefficients of Education*, World Bank Institute, 2001.
- [3] J. A. Weymark, Generalized Gini Indices of Equality of Opportunity, *Journal of Economic Inequality*, Vol. 1, No. 1, pp. 5-24, April, 2003.
- [4] J. Kunegis and J. Preusse, Fairness on the Web: Alternatives to the Power Law, *4th Annual ACM Web Science Conference*, New York, NY, 2012, pp. 175-184.
- [5] D. Braun, Multiple Measurements of U.S. Income Inequality, *Review of Economics and Statistics*, Vol. 70, No. 3, August, 1988, pp. 398-405.
- [6] J. Schwartz and C. Winship, The Welfare Approach to Measuring Inequality, *Sociological Methodology*, Jossey-Bass, 1980, pp. 1-36.
- [7] C. Gini, *Variabilità e Mutabilità: Contributo Allo Studio Delle Distribuzioni e Delle Relazioni Statistiche*, Tipografia di Paolo Cuppini Bologna, 1912.
- [8] L. C. Freeman, Centrality in Social Networks:

- Conceptual Clarification, *Social Networks*, Vol. 1, No. 3, pp. 215-239, 1978-1979.
- [9] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.
- [10] G.-L. Chen, W.-Z. Guo, Y.-Z. Chen and Y.-M. Luo, Mining Influential Communities for Virus Prevention in Cloud-based Network, *Journal of Internet Technology*, Vol. 15, No. 5, pp. 823-834, September, 2014.
- [11] S. A. Myers, A. Sharma, P. Gupta and J. Lin, Information Network or Social Network?: The Structure of the Twitter Follow Graph, *23rd International Conference on World Wide Web Companion*, Seoul, South Korea, 2014, pp. 493-498.
- [12] P. Massa, M. Salvetti and D. Tomasoni, Bowling Alone and Trust Decline in Social Network Sites, *8th IEEE International Conference on Dependable, Autonomic and Secure Computing*, Chengdu, China, 2009, pp. 658-663.
- [13] M. Fire, R. Puzis and Y. Elovici, Link Prediction in Highly Fractional Data Sets, *Handbook of Computational Approaches to Counterterrorism*, Springer, 2012, pp. 283-300.
- [14] G. Palla, I. Farkas, P. Pollner, I. Derényi and T. Vicsek, Directed Network Modules, *New Journal of Physics*, Vol. 9, p. 186, June, 2007.
- [15] L. A. Adamic and N. Glance, The Political Blogosphere and the 2004 U.S. Election: Divided They Blog, *3rd International Workshop on Link Discovery*, Chicago, IL, 2005, pp. 36-43.
- [16] V. Gómez, A. Kaltenbrunner and V. López, Statistical Analysis of the Social Network and Discussion Threads in Slashdot, *17th International Conference on World Wide Web*, Beijing, China, 2008, pp. 645-654.
- [17] J. Leskovec, D. P. Huttenlocher and J. M. Kleinberg, Signed Networks in Social Media, *SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, 2010, pp. 1361-1370.
- [18] J. Leskovec, D. P. Huttenlocher and J. M. Kleinberg, Governance in Social Media: A Case Study of the Wikipedia Promotion Process, *4th International AAAI Conference on Weblogs and Social Media*, Washington, DC, 2010, pp. 98-105.
- [19] B. Viswanath, A. Mislove, M. Cha and K. P. Gummadi, On the Evolution of User Interaction in Facebook, *2nd ACM Workshop on Online Social Networks*, Barcelona, Spain, 2009, pp. 37-42.
- [20] C. Dangalchev, Residual Closeness in Networks, *Physica A: Statistical Mechanics and Its Applications*, Vol. 365, No. 2, pp. 556-564, June, 2006.
- [21] T. F. Liao, Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-based Clustering, *Sociological Methodology*, Vol. 36, No.1, pp. 201-224, December, 2006.
- [22] L. Wittebolle, M. Marzorati, L. Clement, A. Balloi, D. Daffonchio, K. Heylen, P. De Vos, W. Verstraete and N. Boon, Initial Community Evenness Favours Functionality Under Selective Stress, *Nature*, Vol. 458, No. 7238, pp. 623-626, April, 2009.
- [23] World Bank Institute, *Introduction to Poverty Analysis*, 2005, <http://siteresources.worldbank.org/PGLP/Resources/PovertyManual.pdf>
- [24] M. Gupte, P. Shankar, J. Li, S. Muthukrishnan and L. Iftode, Finding Hierarchy in Directed Online Social Networks, *20th International Conference on World Wide Web*, Hyderabad, India, 2011, pp. 557-566.

Biographies



Heeyoung Kwak is PhD candidate in the ECE Dept. at Seoul National University (SNU). She received Bachelor of Science in Electrical & Electronic Engineering from Yonsei University in August 2013. Her research interests are in the field of social network analysis and recommender system.



Joonyoung Kim is PhD candidate in the ECE Dept. at Seoul National University (SNU). He received Bachelor of Science in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST) in February 2012. His research interests are in the field of crowdsourcing and vehicular networks



Yongsu Lim is a postdoctoral researcher in the Department of Computer Science and Engineering of Seoul National University. He received PhD in School of Computing at KAIST. His research interests include graph mining and data stream mining.



Shin-Kap Han is Professor of Sociology and Head of the Department of Sociology at Seoul National University. His areas of interest include: Social Networks; Organizations and Institutions; Careers and Stratification; Quantitative Methods; and Theory Construction. He has published in *Social Networks* and has served on the organizing committee for INSNA.



Kyomin Jung is an associate professor at Seoul National University Electrical and Computer Engineering department. He received his PhD at MIT in 2009, and BSc at Seoul National Univ. in 2003 respectively. His main research areas include social network analysis, and machine learning.

Appendix

Here, we introduce a notion of hierarchy of a network proposed by [24]. They consider a directed acyclic graph (DAG) as a perfect hierarchical structure, and design a measure to see how strong hierarchy a given network has. Concretely, they define hierarchy of a network $G = (V, E)$ as follows (the original definition of hierarchy in their paper is $H'(G) = 1 - \frac{H(G)}{|E|}$).

$$H(G) = \min_{h \in \mathbb{Z}^{|V|}} \left(\sum_{(u,v) \in G} \min(h_u - h_v + 1, 0) \right). \quad (7)$$

They show an integer programming formulation of H , and propose a combinatorial algorithm to exactly minimize H . Note that for a given network G , as $H(G)$ gets smaller, G gets topologically closer to a directed acyclic graph. We call $h^* \in \mathbb{Z}^{|V|}$ corresponding to $H(G)$ a hierarchical structure of a network G .