Centrality, Rational Erdös Number, PageRank:

Assessment in Research Network

Summary

As research society gets bigger and bigger, the demand for evaluating the influence of academic research is growing. Our goal is building assessment models to analyze the influence and impact in research networks and other areas of society. In this paper, we develop three measurement models, namely Mixture of Centrality and Rational Erdös network (MCE), Weighted PaperRank (WPR) and Multigraph-based AuthorRank (MAR).

In Model 1, we combine traditional methods, centrality and rational Erdös number (REN) into MCE to measure influence in co-author networks like Erdös network. MCE is expressed as the ratio of centrality and REN, and it overcomes the defects of them. It is used to estimate influence of authors in Erdös network in Task 1 & 2, and its result is more realistic.

In Model 2, we introduce PaperRank, a variant of Google's PageRank, to measure influence in citation network as given in Task 3. To improve PaperRank's accuracy in relatively small citation network, we combine the number of times cited of a paper into weighting of edges and develop the Weighted PaperRank model. In case of measuring influence of organizations, such as universities, departments or journals, we develop a Multi-granularity WPR model that packs all papers of an organization into one node and then adopts the similar WPR model.

In Model 3, we are to evaluate influence of author in citation network. We put forward a novel influence network, Author-Citation (AC) network, and develop another PageRand based model, AuthorRank (AR). As AuthorRank fails to take parallel edges into account, we improve it by Multigraph-based AuthorRank.

We analyze and compare the results of our models, as well as their strength and weakness. We test our models in Twitter network in Task 4 and analyze the results by their distributions. We also generate a suggestion of using these models to help decision making, as required in Task 5.

Content

1 Introduction

As more and more research papers have been published, researchers and engineers need a persuasive and robust method to help them determine the most influential article or the most authoritative researcher in their field. Authors and co-authorships form a co-author network while papers and citations construct a citation network. This paper aims at developing an effective model to analyze the above-mentioned network, and hopes to extend the model to other field such as social network, if possibly. The following is five tasks we solve in the paper.

- Build up the Erdös1 network. (mentioned in 3.1.1)
- Develop influence measure(s) to determine the most important author in Erdös1 network. (3.1-3.3)
- Build the influence network and determine the most influential paper and researcher in network science. Extend the model to measure influence of an university, a department and a journal. (4-5)
- Apply our algorithm and model to a different kind of network. (7.1)
- Discuss the science, understanding and utility of modeling influence within networks. (7.2)

1.1 Background

Coauthor network and citation network are techniques to determine the influence of academic research. A network is a set of items, vertices or nodes, with connections between them, called edges. In coauthor networks, vertices are researchers and two researchers are considered connected if they coauthored one or more papers together. Researchers in this network are considered to have a strong influence on their coauthors. The influence of each individual can be measured by analyzing the properties of coauthor network. In citation networks, vertices are articles and a directed edge from article A to article B indicates that A cites B. Citation networks are acyclic because papers can only cite other papers that have already been written [1] .

Coauthor network of Paul Erdös is one of the most studied collaboration network, as mentioned in 2014 ICM problem statement. The Erdös number is the distance between a people and Paul Erdös in the Erdös network, and is used as an indicator of one's academic status [2]-[3] . Michael Barr of McGill University has an interesting suggestion for rational Erdös numbers, generalizing the idea that a person who has written p joint papers with Erdös should be assigned Erdös number $1/p$. As many of Erdös coauthor have written more than one joint papers with Erdös, we adopt the rational Erdös numbers for measuring a scholar's influence in this paper.

In graph theory and network analysis, centrality of a vertex measures its relative importance within a graph. There are four main measures of centrality: degree, betweenness, closeness, and eigenvector. Degree centrality is defined as the number of edges incident upon a vertex.

Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. PageRank [4] is an algorithm used by Google Search to rank websites in their search engine results and it is a variant of eigenvector centrality measure. PageRank performs well in page network composed of web pages. It is challenging and worthwhile to extend it to research network.

1.2 Our work

Our innovations in this paper include three evaluation models and a novel influence network. The detail is as follows:

- We .propose a Mixture of Centrality and rational Erdös number (MCE) model to evaluate researcher in coauthor network. MCE considers both the number and quality of co-authorship and overcome deficiencies of its two prototypes.
- We develop a PaperRank based on Google's PageRank to assess research papers within citation network. Then we adjust PaperRank to Weight PaperRank (WPR) exploiting total times cited of papers.
- We propose a novel influence network, author-citation (AC) network, to exploit citation information to reflect researchers' status.
- We successfully extend PageRank to AC network and develop AuthorRank (AR) and Multigraph AuthorRank (MAR) to determine authoritative researchers in a certain field.

In section 3 we introduces centrality, rational Erdös number and their mixture. In section 4 we introduces PageRank and propose PaperRank and Weighted PaperRank. In section 5, we propose AuthorRank and Multigraph based AuthorRank. Section 6 is the experimental result and analysis. Section 7 is a the Twitter network, and find some interesting results. Section 8 includes conclusion and discussion about strength and weakness of our model.

2 Assumptions and Definitions

2.1 Assumptions

- The published time of a paper does not affect measuring influence of the paper. In other words, the influence of a paper is independent to the published time..
- The content of a paper does not affect measuring influence of the paper.

- The journal a paper was published does not affect measuring influence of the paper.
- In Erdös1 network, if two researchers have co-authored papers, we simply assume that they have written only one joint paper due to a lack of related data.
- In Erdös1 network, co-authorship with people whose Erdös number is larger than 1 does not affect measuring influence of Erdös1 authors.
- Given a task to rank papers in a sample set, citations from out of the set could be not considered.
- Given a task to determine the most influential paper or author in a certain field, we assume that the given sample set contains most influential papers and papers written by very authoritative researchers.

2.2 Definition

Variable	Definition
G	a graph
V	a vertex of a graph
E	an edge of a graph
$w(u,v)$	the weight of the edge (u,v)
$C_D(V)$	degree centrality of the vertex V
$rE(V)$	the rational Erdös number of the vertex V
$MCE(V)$	the Mixture of Centrality and rational Erdös number of the vertex V
$PR(V)$	PageRank of vertex V
$L(V)$	the out-degree of the vertex V
B_u	the set of the containing all pages linking to page u
$PPR(u)$	the PaperRank value of a paper u
M	a Matrix
M'	the transpose of matrix M
$WPR(u)$	the weighted PaperRank of paper u
$TC(u)$	the times cited of u
$P(u)$	a set of the papers published by organization u
$AR(u)$	the AuthorRank value of author u
$MAR(u)$	the Multigraph based AuthorRank of author u

Figure 2.1

3 Model 1 : Mixture of Centrality and rational Erdös number

We develop a mixed model, Mixture of Centrality and rational Erdös number (MCE), to discover which author is the most influential in the Erdös1 network as well as in other co-author network. At first, we proposed two traditional measures to solve the task. Then we combine these two measures into our novel MCE model.

3.1 Centrality

We proposed a simple centrality model to evaluate the Erdös1 authors and determine the most influential one in the Erdös1 network. Centrality is a useful and classic concept in graph theory and network analysis. The centrality of a vertex measures its relative importance within a graph. Similarly, each Erdös1 author's importance within the Erdös1 network can be measured by his centrality. To discover the relative influence among individuals in the Erdös1 network, we firstly construct the network. Secondly, we calculate the centrality value of each Erdös1 author and sort them according to centrality. Person with higher centrality is more important. The following is the details.

3.1.1 Build up the Erdös1 network

To build up the Erdös1 network, we employ a vertex to represent an Erdös1 author and an undirected edge to represent a co-authorship between two authors. That is to say, supposing that vertex x represents author A and vertex y represents author B, an edge exists between x and y if and only if A has co-authored some research papers with B. Take notice that A has co-authored papers with B if and only if B has co-authored papers with A. Edges in Erdös1 network are undirected as co-authorship is mutual. Thus the Erdös1 network is an undirected graph.

3.1.2 Calculate the Centrality

Actually centrality of a vertex consists of multiple attributes, including degree centrality, closeness centrality, betweenness centrality, Eigenvector centrality, etc. We choose some of them to evaluate the influence of Erdös1 authors.

Degree centrality is historically first and conceptually simplest. It is defined as the number of edges incident upon a vertex. Formally, the degree centrality of a vertex v, for a given graph $G:=(V, E)$ with $|V|$ vertices and $|E|$ edges, is defined as $C_D(v) = \deg(v)$. In the Erdös1 network, suppose that a vertex has more edges linking to other vertex. The Erdös1 author the mentioned vertex represents has more co-authors and shows more influence to some extent. Thus we just need to find out the author with the greatest degree centrality and he is the most influential one in Erdös1 network.

In short, we mainly consider the degree centrality of Erdös1 authors to determine the member that is most important in the network or has a closet academic relationship with Erdös.

3.2 Rational Erdös Number

Although the above-mentioned centrality model is simple and useful, it has a serious deficiency that it does not consider how close the academic relationship between Erdös and different collaborators is. Consider an author A that has co-authored 50 papers with Erdös and an author B that has co-authored 5 papers with Erdös. Even if A and B have the same degree centrality, their influence in the Erdös network should not be the same. Apparently, A is more influential than B as A co-authored more papers and shows a tougher academic relation with Erdös.

To overcome the deficiency in centrality, we propose a rational Erdös number model based Michael Barr's paper. The following is a concise introduction to Erdös number. By definition, Erdös had Erdös number zero, his direct collaborators had Erdös number one, a person who do not collaborated with Erdös, but had collaborated with one of his collaborators has Erdös number two and so on. Actually, an author's Erdös number equals the length of the shortest path between the author and Erdös in a complete Erdös1 network. Length of a path is the number of edges in the path. A complete Erdös1 network is the Erdös1 network with Erdös and his co-authorship. That is to say, in the complete Erdös1 network, there exists a vertex denoting Erdös and the vertex has an edge incident upon each other vertex in the network. If an author has smaller Erdös number, he is academically closer to Erdös and shows more influence in Erdös1 network.

To consider the number of papers collaborated between Erdös and his collaborator, Michael Barr developed rational Erdös numbers. A person who has written five joint papers with Erdös is assigned the rational Erdös number 1/5. If a person has written two joint papers with someone whose Erdös number is 1, then his Erdös number ought to be 3/2 (1+1/2=3/2). If you have written two joint papers with someone who has written three joint papers with Erdös, then your rational Erdös number should be 5/6 (1/3+1/2=5/6) and so on. In the following, we denote rational Erdös number by rE.

In a way, the rational Erdös number model turns the original un-weighted Erdös1 network into a weighted graph. In the weighted graph, each edge between two vertices is assigned a weight that equals $1/N$. N is the number of papers written by the Erdös1 authors that the two vertices denote. Each author's rE equals the shortest distance between Erdös and him. Similar to Erdös number model, person who has smaller rE is more influential in the Erdös1 network.

Limited by our dataset, we could learn the number of papers co-authored by Erdös and his collaborators but we could not know the number of joint papers written by two Erdös1 authors. We just know whether two Erdös1 authors have collaborated. We simply assume that if two Erdös1 authors have collaborated they have written one joint paper. In a way, if a person has co-authored more papers with Erdös, his rE is smaller and he is more important in the Erdös network.

3.3 Mixture of Centrality and rational Erdös number

The rational Erdös number model considers that different Erdös authors may have different collaborative extent with Erdös, but it does not emphasize the co-authorship between the Erdös1 authors. Compared to the rational Erdös number model, the centrality model emphasizes the collaborations among the Erdös1 authors but does not employ the number of papers between Erdös and his collaborators. In short, the advantages of the centrality model and the rational Erdös number model are complementary. That inspires us to develop a Mixture of Centrality and rational Erdös number (MCE) model.

We use *MCE* to denote the value of an author measure by MCE model. $MCE(A)$ is defined as author *A*'s *MCE* and is computed as

$$MCE(A) = \frac{C_D(A)}{rE(A)} \tag{1}$$

where $C_D(A)$ is the degree centrality of author *A* and $rE(A)$ is the rational Erdös number of *A*. We adopt the ratio of $C_D(A)$ and $rE(A)$ as *A*'s *MCE* because person with higher degree centrality is more influential and because author with lower *rE* has a closer academically relationship to Erdös.

4 Model 2 : Weighted PaperRank

In this section, we first introduce the PageRank algorithm used in link network. Then we propose a new algorithm based on PageRank called PaperRank, and exploit it in measuring influence in citation network. We then further improve this algorithm by altering weight of edges with the total number of citation of a paper.

4.1 PageRank

PageRank [4] is a link network analysis algorithm and it assigns a numerical weighting to each node in the network. In link network, nodes are webpages and there is a directed edge from webpage A to webpage B if there is a hyperlink from A to B. The numerical weight assigned to any given node *p* is referred as the PageRank of *p* and denoted by $PR(p)$. In general case, PageRank value of a node *u* can be expressed as:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)} \tag{2}$$

where B_u is the set containing all pages linking to page *u*, $L(v)$ is the number of links from page *v*. The PageRank value of any given page *u* is dependent on the PageRank value of each page *v* contained in B_u.

4.2 PaperRank

We modify PageRank algorithm to analyze the influence and impact in citation network. Citation networks are similar to link networks. They are both modeled as directed graph and the edges are of similar meaning. In link network, such as world-wide-web, the links from one page to another can be considered as a kind of citation. However, citation network is acyclic or rarely acyclic, because a paper can only cite others papers that have already been written. The

PaperRank value of a paper p can expressed as

$$PPR(p) = \sum_{q \in B_p} \frac{PPR(q)}{L(q)} \tag{3}$$

where B_p is the set containing all papers which cite paper p, $L(q)$ is the number of citations from paper q. Unlike PageRank, we set the damping factor d to 1, and therefore it is omitted in the equation. It is academic standard that the author should cite all the papers he referenced. So, the random behavior in the damping factor model does not exist and we set d to 1. Setting d to 1 will not lead to spider trap problem as in WWW will not occur simply because a paper cannot cite itself. However, dead end problem still exists in PaperRank. We simply assume that a paper that does not cite any other papers cites all the other papers. Written in matrix notation, PaperRank of a citation network can be expressed as a vector

$$\mathbf{R}' = \left[PPR(p_1)\ PPR(p_2) \cdots PPR(p_N)\right]^T \quad , \tag{4}$$

where $p_1, p_2, \ldots, p_N$ are the papers under consideration. The matrix M is defined as

$$M = \begin{cases} \dfrac{1}{L(p_j)}, & if\ j\ cites\ i \\ 0, & otherwise \end{cases} \tag{5}$$

and the equation is as follows: $R' = MR'$ Iterative form of this equations: $R'(t+1) = MR'(t)$

where $R'_i(\mathrm{t}) = PPR(p_i;t)$ and t is the number of iterations. At t=0, an initial probability distribution is assumed, usually $PPR(p_i;t) = 1/N$

The computation ends when for some small ε, $|R'(t+1) - R'(t)| < \varepsilon$. It is easy to prove that $R'(t)$ eventually converge to the eigenvector of M. In this point of view, PaperRank is a variant of the eigenvector centrality measure of the citation network

4.3 Weighted PaperRank

In this part we develop an extension of PaperRank, weighted PaperRank (WPR) model, exploiting total times cited of papers. PaperRank has a deficiency and may make mistakes in determining relative importance among papers. When we use PaperRank algorithm, we usually apply it to a small set of dozens of nice papers. The underlying assumption is that the most influential papers have been included in the set. As PaperRank only exploits citation relationship between papers in the set and it may ignore the citations and supports from papers out of the set. That could lead to unacceptable and obvious mistakes. Assume that paper A and B are in the set. The total times cited of paper A is huge while that of paper B is very small. In general A is more influential than B. However, if we ignore the citations from papers out of the set (actually that is how PaperRank do) and just count the citations from papers in the set, the times cited of paper A may even be less than that of paper B, which could result in that B ranks higher than A.

To enhance the robustness of the algorithm, we consider total times cited of papers to avoid judgment mistakes. In detail, if a paper cites several other papers, it does not simply give these papers the same support but refer to total times cited of these papers. Papers with higher total times cited are received more support. PaperRank deals with citation network consisting of nodes representing papers and directed edges representing citation relation. In PaperRank, citation network is an un-weighted graph. Now we assign edges $e(u,v)$ a weight according to the extent of support that v receives from u and extend the citation network to a weighted one. Thus we name the extension algorithm weighted PaperRank (WPR). We denote weighted PaperRank value by *WPR* later.

The following is the details of WPR algorithm. We denote the *WPR* of paper u by $WPR(u)$. Firstly, we initialize the *WPR* of each paper in the citation network by $WPR(u)=1/|V|$ where $WPR(u)$ denotes the *WPR* of u, V denotes the set of papers, $|V|$ denotes the size of V. Secondly, we update the *WPR* of each paper u by

$$WPR(u) = \sum_{(v,u)\in E} \frac{WPR(v)\cdot TC(\mathrm{u})}{\sum_{(v,x)\in E} TC(x)} \tag{6}$$

where $WPR(u)$ is the *WPR* of u, $WPR(v)$ is the *WPR* of v, $TC(u)$ is the times cited of u, (v,u) denotes a directed edge from v to u, E is the set containing all edges. After we have updated *WPR* of all papers, we begin next iteration to update them again until *WPR* of each paper converges (do not change any more). At last, $WPR(u)$ holds the *MAR* score of u and the paper with highest *WPR* is regarded as the most authoritative and influential paper in the citation network.

4.4 Multi-granularity Weighted PaperRank

The WPR algorithm is effective in measuring influence and impact of a paper in citation network. How can we measure the influence and impact of a university, department or a journal based on the citation network? The basic assumption is that by measuring the influence of papers written or published by an organization, we can determine its influence.

We revise our WPR model into a Multi-granularity WPR (MWPR). First, we build the multi-granularity citation network, denoted as $G_1=(V_1,E_1)$. We denote the citation network in previous section as A node in this network is a paper or an organization. For any organization node u, there is a directed edge from u to node v if there exists a paper of u that cites v or a paper of v (if v is an organization). For any paper node u, there is a directed edge from u to node v if u cites v or a paper of v (if v is an organization). To achieve the goal of building G_1, we have to dataset of papers written or published by an organization u, which is denoted as follows:

$$P(u) = \begin{cases} \{\text{papers of } u\}, & \text{if u is organization} \\ \{u\}, & \text{if u is paper} \end{cases} \tag{7}$$

Next, we adopt the similar method of WPR to to multi-granularity citation network. The difference is that the weight of edges in this network is the number of citations between the two nodes. We denote former citation network as $G_2=(V_2,E_2)$. The weight of each edge (u,v) in G_1 is expressed as follows:

$$w(u,v) = \sum_{e\in\{(p,q)|E_2, p\in P(u), q\in P(v)\}} w(e) \tag{8}$$

Finally, by applying the WPR algorithm to G_1, we can get the MWPR value of the organizations we want to evaluate.

5 Model 3 : Multigraph Based AuthorRank

We develop a Multigraph Based AuthorRank (MAR) model to determine the most influential and important researcher in network science or some other research fields. Our second model, WPR, is used to discover the relative influence among multiple papers in network science. WPR deals with citation network. Each node in citation network represents an article. Thus WPR could not help find out the most important author in network science and we need a new model. That is the reason we propose MAR.

MAR is also a PageRank based model similar to WPR. At first, we manage to construct an author-citation network. Then we apply PageRank method to it and develop the AuthorRank method. Considering the distinction between author-citation network and link network that PageRank was initially used in, we adjust AuthorRank to a more desirable model, Multigraph based AuthorRank, which shows the relative importance among authors in the same field well.

5.1 AuthorRank

5.1.1 Build up an Author-Citation Network

We propose a novel author-citation network (AC network) as citation network could not reflect attributes of authors. A vertex in AC network represents an author. If in one of author *A*'s paper, *A* cited one article of author *B*'s, a directed edge from *A* to *B* is added to the AC network. In this way, we could construct a directed graph with nodes denoting authors and edges reflecting citation relation among authors, namely an author-citation network. Take notice that there only exists one directed edge from author *A* to author *B* even if *A* has cited *B*'s papers many times for simplicity. However, we allow a directed edge from *B* to *A* to be added in this case. Because the condition that two researchers have cited each other's papers is usual and reasonable.

5.1.2 PageRank in Author-Citation Network

PageRank is an algorithm used by Google Search to rank websites in their search engine results. It works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites [5] . Form another point of view, PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. In PageRank, a page network consists of nodes representing pages and directed edges representing links. Take notice that the page network has no multiple edges (also called “parallel edges”) in general because multiple links to the same page may not add the probability of click to the page.

Similar to PageRank, we assume that for more authoritative researchers their papers are likely to be cited from other papers. If you cite someone’s papers in your article, you may support his opinions or recognize his authority to some extent. Thus it is reasonable to apply PageRank to an Author-Citation network. That is AuthorRank, a simple extension of PageRank to AC network.(We use *AR* to denote an AuthorRank value later.)

The following is the details of AuthorRank algorithm. We denote the AuthorRank value of author u by $AR(u)$. Firstly, we initialize the AuthorRank score of each author in the AC network by $AR(u)=1/|V|$ where $AR(u)$ denotes the AuthorRank value of u, V denotes the set of authors, $|V|$ denotes the size of V. Secondly, we update the *AR* of each author u by

$$AR(u) = \sum_{v \in B_u} \frac{AR(v)}{L(v)} \tag{9}$$

where $AR(u)$ is the AuthorRank value of u, $AR(v)$ is the *AR* of v, B_u is the set containing all authors have cited u’s papers, $L(v)$ denotes the number of directed edges from v. After we have updated all authors’ *AR*, we begin next iteration to update them again until each author’s *AR* converges (do not change any more). At last, $AR(u)$ holds the AuthorRank score of u and the author with highest *AR* is regarded as the most authoritative and influential researcher in the AC network.

To determine the most influential author in some fields such as network science, a proper set of dozens of papers about the target field should be formed carefully. Papers satisfying the standards listed below are desirable to choose in the set.

- Papers are related to or concern the same field.
- Papers are fundamental, very authoritative or highly cited.
- Papers have multiple citation relationships among them.
- One of authors may be most authoritative or influential.

Given a proper set of papers, we build up an AC network according to the set and run the AuthorRank algorithm upon the network. Then we could discover the most influential and important author in the network.

5.2 Multigraph Based AuthorRank

5.2.1 Deficiency in AuthorRank

In order to extend PageRank to AuthorRank easily, AC network avoids parallel edges as page network also avoids them. However, avoiding parallel edges is not so reasonable and thus limits the AuthorRank algorithm. Assume that author *A*, *B* and *C* are in the AC network. Suppose that *A* has cited *B*'s many papers for many times while *A* has cited *C*'s paper for one time. It is very likely that *A* regards *B* more influential than *C*, which is not reflected in AC network. There exists only one un-weighted edge from *A* to *B* even if *A* has cited *B*'s papers for many times.

5.2.2 AC network with Parallel Edges

To overcome the deficiency in AuthorRank, we allow parallel edges and adjust AC network to a multigraph. If author *A* has cited author *B*'s paper for *T* times, *T* edges from *A* to *B* are added to the network. Then we develop a novel Multigraph based AuthorRank (MAR) model. We denote the MAR value by *MAR* later.

The following is the details of MAR algorithm. We denote the *MAR* of author *u* by $MAR(u)$. Firstly, we initialize the *MAR* of each author in the AC network by $MAR(u)=1/|V|$ where $MAR(u)$ denotes the *MAR* of *u*, *V* denotes the set of authors, $|V|$ denotes the size of *V*. Secondly, we update the *MAR* of each author *u* by

$$MAR(u) = \sum_{v \in B_u} \frac{MAR(v) \cdot L_u(v)}{L(v)} \tag{11}$$

where $MAR(u)$ is the *MAR* of *u*, $MAR(v)$ is the *MAR* of *v*, B_u is the set containing all authors have cited *u*'s papers, $L(v)$ denotes the number of directed edges from *v*, $L_u(v)$ denotes the number of edges from *v* to *u*. After we have updated all authors' *MAR*, we begin next iteration to update them again until each author's *MAR* converges (do not change any more). At last, $MAR(u)$ holds the *MAR* score of *u* and the author with highest *MAR* is regarded as the most authoritative and influential researcher in the AC network.

6 Results and Analysis

Figure 6.1 is the Erdös1 network consisting of 511 nodes. Figure 6.2 shows the degree distribution of the Erdös1 network. As Figure 6.2 shows, the number of vertices whose degree is less than or equals 5 is 278, and the number of vertices whose degree is less than 10 and larger than 5 is 134. The maximum and minimum of the degree are 52 and 0 respectively. The average of degree in the Erdös network is 6.41487. The Erdös1 network is a undirected graph because co-authorship is mutual and bidirectional.

Table 6.1 shows ten most influential authors measured by degree centrality. ALON and HARARY have written 5 and 2 papers with Erdös respectively but they have more than 40 Erdös1 co-authors, which make them rank high in degree centrality model. We conclude that degree centrality focus on the number of co-authorships instead of the quality and closeness of co-authorships. Table 6.2 shows ten most influential authors measured by rational Erdös number model. We discover that author that ranks higher has smaller rational Erdös number and more papers collaborated with Erdös. Thus rational Erdös number model emphasize the quality of co-authorship. Table 6.3 shows the most important authors in Erdös1 network in MCE model. Authors who rank higher have more joint papers and more co-authors relatively. In other words, our MCE model keeps a balance between the number and quality of co-authorships and is more acceptable, reasonable and desirable.

Table 6.4 and 6.5 are 16 papers respectively ranked by PaperRank and Weighted PaperRank. Cited is the times cited without considering citations out of the set and Citing is the times citing with the set. Two papers with huge Total Cited 10616 and 132250 rank 6th and 11th respectively, which shows PaperRank lacks exploiting Total Cited. Our WPR considers citations within and outside the given paper set and shows more robustness in ranking. Thus "10616" and "132250" raise from 6th and 11th to 4th and 5th.

Table 6.6 and 6.7 are the ranking of 15 authors in the given set by AuthorRank and Multigraph based AuthorRank respectively. In table 6.6, Watts,D with more papers get a score as the same as his co-author Strogatz,S, which is unfair for authors who have multiple papers cited. MAR considers multiple citations and make Watts,D, Borgatti,S and Dodds,P raise in Table 6.7.

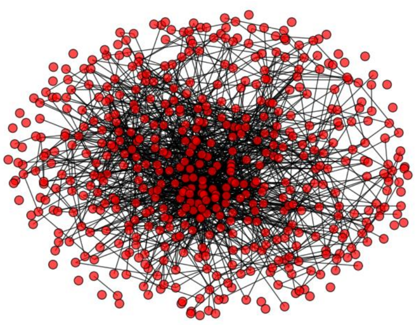

Figure 6.1 the Erdös1 network

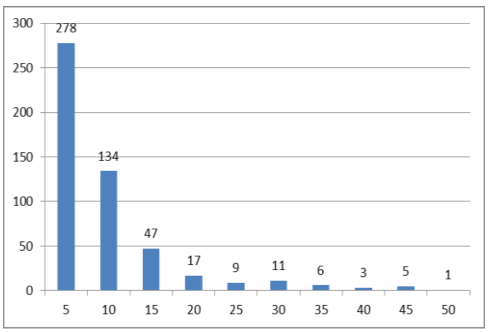

Figure 6.2 degree distribution of the Erdös1 network

Table 6.1 Most influential authors in degree centrality model

Rank	Author	C_D	Papers	Coauthors
1	ALON, NOGA M.	52	5	52
2	HARARY, FRANK*	44	2	44
3	GRAHAM, RONALD LEWIS	44	28	44
4	BOLLOBAS, BELA	43	18	43
5	RODL, VOJTECH	43	11	43
6	TUZA, ZSOLT	40	11	40
7	FUREDI, ZOLTAN	39	10	39
8	SOS, VERA TURAN	38	35	38
9	SPENCER, JOEL HAROLD	35	23	35
10	PACH, JANOS	32	21	32

Table 6.2 Most influential authors in rational Erdös number model

Rank	Author	*rE*	Papers	Coauthors
1	SARKOZY, ANDRAS	0.016129	62	25
2	HAJNAL, ANDRAS	0.017857	56	30
3	FAUDREE, RALPH JASPER, JR.	0.02	50	30
4	SCHELP, RICHARD H.	0.02381	42	26
5	ROUSSEAU, CECIL CLYDE	0.028571	35	14
6	SOS, VERA TURAN	0.028571	35	38
7	RENYI, ALFRED A.*	0.03125	32	12
8	TURAN, PAL*	0.033333	30	12
9	SZEMEREDI, ENDRE	0.034483	29	29
10	GRAHAM, RONALD LEWIS	0.035714	28	44

Table 6.3 Most influential authors in MCE model

Rank	Author	*MCE*	Papers	Coauthors
1	HAJNAL, ANDRAS	1680	56	30
2	SARKOZY, ANDRAS	1550	62	25
3	FAUDREE, RALPH JASPER, JR.	1500	50	30
4	SOS, VERA TURAN	1330	35	38
5	GRAHAM, RONALD LEWIS	1232	28	44
6	SCHELP, RICHARD H.	1092	42	26
7	SZEMEREDI, ENDRE	841	29	29
8	SPENCER, JOEL HAROLD	805	23	35
9	BOLLOBAS, BELA	774	18	43
10	PACH, JANOS	672	21	32

Table 6.4 PaperRank of 16 papers

Rank	PaperRank	Title	Total Cited	Cited	Citing
1	0.243822	Collective dynamics of `small-world' networks	21688	6	0
2	0.123558	Power and Centrality: A family of measures	1965	3	0
3	0.0844316	Emergence of scaling in random networks	18843	2	1
4	0.0556013	Identity and search in social networks	835	2	2
5	0.0514827	Models of core/periphery structures	763	1	1
6	0.0494234	The structure and function of complex networks	10616	1	7
7	0.0469522	On Random Graphs	820	2	0
8	0.0469522	Navigation in a small world	1246	2	1
9	0.0469522	Scientific collaboration networks: II	1523	2	0
10	0.0469522	The structure of scientific collaboration networks	2748	2	1
11	0.0391269	Statistical mechanics of complex networks	132250	1	5
12	0.0329489	Identifying sets of key players in a network.	304	0	0
13	0.0329489	What is your Ramsey number?	16	0	0
14	0.0329489	Networks, influence, and public opinion formation	680	0	2
15	0.0329489	Statistical models for social networks	33	0	1
16	0.0329489	Social network thresholds in the diffusion of innovations	498	0	1

Table 6.5 Weighted PaperRank of 16 papers

Rank	PaperRank	Title	Total Cited	Cited	Citing
1	0.293109	Collective dynamics of `small-world' networks	21688	6	0
2	0.105151	A family of measures	1965	3	0
3	0.0848359	Emergence of scaling in random networks	18843	2	1
4	0.0652313	The structure and function of complex networks	10616	1	7
5	0.0534599	Statistical mechanics of complex networks	13250	1	5
6	0.0431589	The structure of scientific collaboration networks	2748	2	1
7	0.0390091	Scientific collaboration networks	1523	2	0
8	0.0380707	Navigation in a small world.	1246	2	1
9	0.0375539	Identity and search in social networks	835	2	2
10	0.0366276	On Random Graphs	820	2	0
11	0.0345437	Models of core/periphery structures. Social Networks	763	1	1
12	0.0338498	Identifying sets of key players in a network	304	0	0
13	0.0338498	Models of core/periphery structures. Social Networks	16	0	0
14	0.0338498	Networks, influence, and public opinion formation	680	0	2
15	0.0338498	Statistical models for social networks.	33	0	1
16	0.0338498	Social network thresholds in the diffusion of innovations	498	0	1

Table 6.6 AuthorRank

Rank	Author Name	*AR*	Article Numbers	Times Cited
1	Bonacich, P.F.	0.168569	1	1965
2	Strogatz, S.	0.104196	1	21688
3	Watts, D.	0.104196	3	680;835;21688
4	Newman, M.	0.0760386	4	1523;2748;10616;835
5	Albert, R.	0.0673521	2	13250;18843
6	Barabási, A-L.	0.0673521	2	13250;18843
7	Erdös, P.	0.0563156	1	820
8	Kleinberg, J.	0.0563156	1	1246
9	Rényi, A.	0.0563156	1	820
10	Borgatti, S.	0.0535879	2	304;763
11	Dodds, P.	0.0535879	2	680;835
12	Everett, M.	0.0535879	1	763
13	Graham, R.	0.0275283	1	16
14	Snijders, T.	0.0275283	1	33
15	Valente, T.	0.0275283	1	498

Table 6.7 Multigraph Based AuthorRank

Rank	Author Name	*AR*	Article Numbers	Times Cited
1	Bonacich, P.F.	0.153798	1	1965
2	Watts, D.	0.134959	3	680;835;21688
3	Strogatz, S.	0.111468	1	21688
4	Newman, M.	0.107175	4	1523;2748;10616;835
5	Albert, R.	0.0667367	2	13250;18843
6	Barabási, A-L.	0.0667367	2	13250;18843
7	Borgatti, S.	0.0489192	2	304;763
8	Dodds, P.	0.0489192	2	680;835
9	Everett, M.	0.0489192	1	763
10	Erdös, P.	0.0453618	1	820
11	Kleinberg, J.	0.0453618	1	1246
12	Rényi, A.	0.0453618	1	820
13	Graham, R.	0.0254278	1	16
14	Snijders, T.	0.0254278	1	33
15	Valente, T.	0.0254278	1	498

Article Numbers is the number of papers in the given set written by the author.

Times Cited is the total times cited of the author’s papers.

AR is the AuthorRank value.

7 Application and Discussion

7.1 Application in Twitter network

In this part, we exploit the centrality model and PaperRank model in Twitter network analyze our result. Twitter, according to twitter.com itself, is a “Social networking and microblogging service utilising instant messaging, SMS or a web interface.” As a social network, Twitter revolves around the principle of followers. When you choose to follow another Twitter user, that user's tweets appear in reverse chronological order on your main Twitter page. We believe that a person has influence on his/her followers, and the more followers he/she has, the more influential he/she is. The nodes in Twitter network are people, there is a directed edge from person A to person B if A follows B.

We found the a dataset of Twitter network from Stanford Network Analysis Platform[1]. The Twitter network data locates in the Social Network catalog of Stanford Large Network Dataset Collection, and can be downloaded with the link

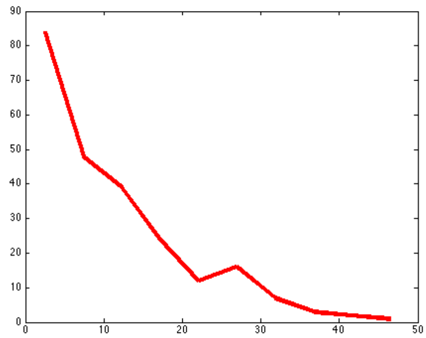

Figure 7.1 Degree centrality distribution

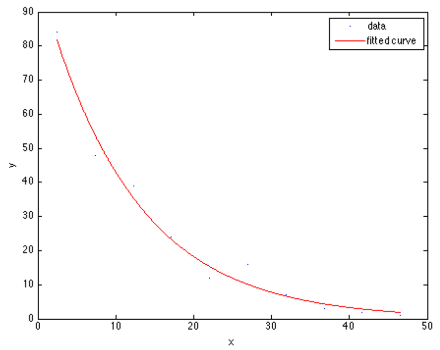

Figure 7.2 Exponential fit curve of distribution

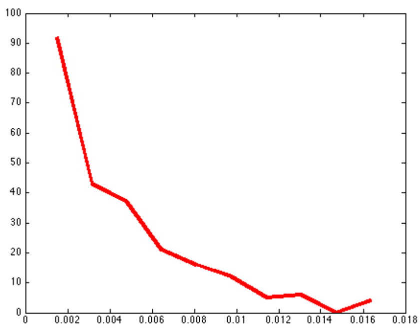

Figure 7.3 PaperRank distribution

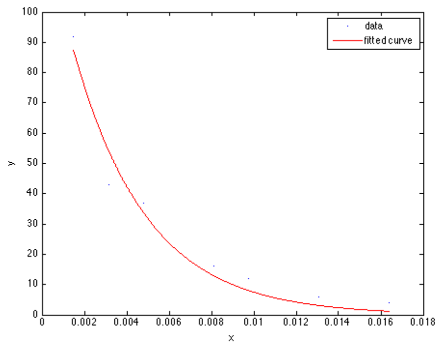

Figure 7.4 Exponential fit curve of distribution

http://snap.stanford.edu/data/twitter.tar.gz. The dataset includes node features (profiles), circles, and ego networks. There are more than 800 ego-networks in this dataset. Ego networks consist of a focal node ("ego") and the nodes to whom ego is directly connected to (these are called "alters") plus the ties, if any, among the alters[2].

According to the definition of ego network, the Erdös1 network we build in section two is a kind of ego network. We choose the ego networks in the file 12831.edges and apply our models in this network. There are total 236 nodes and 2478 edges in this network.

Centrality Model

We calculate the degree centrality of each node, below Figure 7.1 shows the distribution of degree centrality and Figure 7.2 shows its exponential fit curve. The exponential fit curve is $y=ae^{bx}$, x represent index of nodes, and y the degree centrality. a= 108, b= -0.08526.

PaperRank Model

We calculate the PaperRank of each node, below Figure 7.3 shows the distribution of PaperRank and Figure 7.4 shows its exponential fit curve. The exponential fit curve is $y=ae^{bx}$, where a= 134.2, b= -288.

7.2 Discussion

The models we devise can be used in various networks. As show in Twitter network, our models can be used to measure the influence and impact of social network and it is expected to suit other networks. The PaperRank and AuthorRank models are a variant of PageRank, and the basic idea behind them is that one node's influence on others depends on both its current rank and its out-degree. The higher rank a node has the more influence it will distribute to its neighbors. It can be used in a variety of situations, both directed and undirected networks. However, the weakness and limitation is its extendability. The result of this algorithm is mostly determined by the adjacency matrix, which means we can only extend this algorithm by altering the weight of edges.

Once we measure the influence in the network, we can use it to help individuals, organizations, nations and society in making decisions, improve relationships and conduct business. For instance, if you are a researcher who wants to boost your influence in your area, according to AuthoRank model, you should look for people who are of high influence and a relatively small number of coauthor. Same method can be used to help making decisions of organizations, nations or society, using Multi-granularity WPR.

2 http://www.analytictech.com/networks/egonet.htm

8 Conclusions

MCE model shows more reasonability than its two prototypes, degree centrality and rational Erdös number. WPR is a successful trial to extend famous PageRank to citation network. AC network is a novel influence network proposed in this paper. MAR is another desirable extension of PageRank. MAR performs well in discovering authoritative researchers within AC network.

8.1 Strengths and Weaknesses

The advantage of MCE is that it keeps a good balance between the number and quality of co-authorships. MCE considers both of them and thus performs well. The strength of AC network is that it reflects information of both authors and papers and it exploits citation among papers to measure the author.

WPR and MAR are used to measure papers and authors in a certain field. In general they are applied to a sample set of papers. The sample set is supposed to contain most influential papers and papers written by authoritative researchers. WPR and MAR are sensitive to the sample set. If the set is not chosen properly, it may impact performance of WPR and MAR.

8.2 Future Work

Much research needs to do. Our future work is listed below:

- Older papers may get more chance to receive citations. It is useful to develop a model that could measure papers published at difference time.
- PageRank has a lot of variants. Have a try to extend them to co-author network, citation network and author-citation network.
- Different from page network (consist of web pages and links), citation network is acyclic. Discover a fast or simple variant of PaperRank.

9 References

[1]. Newman M E J. The structure and function of complex networks[J]. SIAM review, 2003, 45(2): 167-256.

[2]. Goffman C. And what is your Erdös number?[J]. The American Mathematical Monthly, 1969, 76(7): 791-791.

[3]. Odda T. ON PROPERTIES OF A WELL‐KNOWN GRAPH OR WHAT IS YOUR RAMSEY NUMBER?[J]. Annals of the New York Academy of Sciences, 1979, 328(1): 166-172.

[4]. Page L, Brin S, Motwani R, et al. The PageRank citation ranking: bringing order to the web[J]. 1999.

[5]. http://www.google.com/competition/howgooglesearchworks.html

[6]. http://en.wikipedia.org/wiki/PageRank