

Complex Network based Recommender System

John Kingsley Arthur
Jiangsu University
School of Computer Science &
Comm. Engineering
Zhenjiang, China

Ronky Francis Doh
Jiangsu University
School of Computer Science &
Comm. Engineering
Zhenjiang, China

Eric Appiah Mantey
Jiangsu University
School of Computer Science &
Comm. Engineering
Zhenjiang, China

Jeremiah Osei-Kwakye
Jiangsu University
School of Computer Science &
Comm. Engineering
Zhenjiang, China

ABSTRACT

The structural abstraction of Recommender Systems has a high resemblance as that of a Complex Network(CN). The underlining principles of the different types of Recommender Systems, such as Collaborative Filtering, Content-Based, Knowledge-Based, Utility-Based, and Hybrid are similar to the attributes and theories of Complex Networks. Considering the enormity of computation in Recommender Systems logic formulation and it's a structural similarity with Complex Networks, the research work seeks to highlight attributes of Complex Networks and Recommender Systems to suggest the applicability of theories, and principles in CN's that can be considered in the construction of Recommender Systems. The research further elaborates on matrices for checking the correctness of the use of these theories.

General Terms

Internet, Node, Recommender Systems

Keywords

Complex Networks, Recommender Systems, Collaborative Filtering, Content-Based RS, Hybrid Recommender Systems

1. INTRODUCTION

All around us, a kind of Network is formed. The interactions between us(people) create that natural phenomenon of a social network. A person's close relations expand, as friends become friends with others based on commonalities between them. These structural relations are exhibited not only amongst humans but also in some other biological systems such as protein-to-protein network, food webs, gene-regulatory network, and many others. Networks can be tangible objects in the Euclidean space, such as electric power grids, the Internet, highways or subway systems, and neural networks[1]. In the last ten (10) years, a vast amount of research works have routed massive focus on the study of Complex Networks. Complex Networks are systems whose structure is irregular, complex and dynamically evolving in time, and also having its main focus heading from the analysis of small networks to that of systems with a large number of nodes, and with a rehabilitated devotion to the properties of networks of dynamical units [1]. Complex networks feature patterns of connections between their elements that are neither purely regular nor purely random. The Internet, the World Wide Web, online social networks such as WeChat, Facebook, LinkedIn, Instagram, Twitter, Whatsapp are representations of real-world examples.

Complex networks have been used to solve challenging problems in the areas of transportation, computer system breaches, medicine to understand chemical compositions and advert effects, the formation of music communities, pattern recognition, and many more other areas[1][2] and have achieved some level of success.

The principles and theories of Complex Networks apply to an intelligent computational area such as recommendation. To understand the scenario of recommendation very well, the antique approach of purchasing clothes is used. A customer has to enter a shop and scout around to find what he/she wants or talk to the attendant. The shop attendant can suggest to you what to select if he/she knows your expectations or can also choose to do guesswork by your immediate appearance. However, for these recent years, especially with the influx of e-commerce sites, there is a huge number of online shops to buy clothes from. The information heap is so huge, so much so that customers are unable to consume it all. It is in this direction that recommendation systems are unveiled to serve as personalized choice assistant to suggest items to users based on some collected data.

A Recommendation System is an intelligent system that is capable of suggesting items of interest to a person based on an underlining algorithm that learns from data [2]-[5]. Recommender systems are utilized in a variety of areas including movies, games, music, news, books, dating sites, research articles, social tags, and other kinds of products. There are also recommender systems for experts, collaborators, jokes, restaurants, garments, financial services, life insurance, romantic partners (online dating), and Twitter pages. A very common example as well is seen with Amazon on how pages are personalized for individual users based on preferences. Likewise, is that of Youtube. It enhances search on the web and makes it easier [5].

When constructing an RS basing on either Collaborative Filtering, Content-Based, Knowledge-Based, and Hybrid RS techniques, a logical complex network is created between the nodes representing either items or customers. Complex Network has properties such Small the Small World Effect, Bipartite Graph formation, Cyclic co-efficient, and transitivity properties can be used as the basis to tell the interrelation between nodes on the network and by that make a recommendation to the user. Upon this concept we coin the term "Complex Network Based Recommender System" for all RSs built on the principles of Complex Networks.

The rest of the document is divided into six(6) sections. In section 2, the attributes of complex networks are discussed. In section 3, the different kinds of Recommender systems and techniques used in their logical make up are elaborated. In section 4, some algorithm measuring techniques are looked at. Section 5 gives a brief introduction of how to build results from a recommender system is treated. Lastly, section 6 the conclusion and direction for future work is presented.

2. STRUCTURAL ATTRIBUTES OF COMPLEX NETWORKS

As established earlier, a complex network has many nodes that interact with each in a defined manner. This structure defines some corresponding qualities such as the Scale-free networks and networks with community structure. In this session, some relevant attributes of Complex Networks for studying Recommender Systems are highlighted.

2.1 Small World Effect

Several real-world networks show what is termed as the small world effect property where most of the vertices or nodes are reached from others through a smaller number of edges [6] as shown in fig. 1 below. In the small world, nodes form acquaintances with themselves where, for example, the friends of node A have some forms of friendship with the friends of Node B.

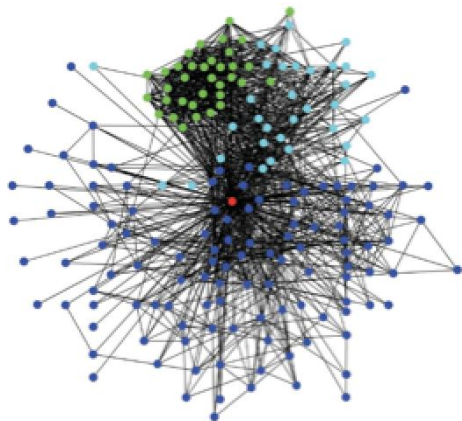


Fig 1: Small World Network Structure Source: Adapted from [8]

In the area of social networks, the normal separation starting with one node then onto the next is little contrasted with the networks measure, known as the Small World Effect. The size of a network is generally estimated by identifying the number of vertexes(nodes) in the social network[12]. This Small World Effect is similar to the logic behind Collaborative Filtering, where items are recommended to the current user based on their common likes.

2.2 Cyclic Coefficient

Kim and Kim[6] defined a coefficient to measure how cyclic a network is. The local cyclic coefficient of a vertex i is defined as the average of the inverse of the sizes of the smallest cycles formed by vertex i and its neighbors,

$$\theta_i = \frac{2}{k_i(k_i - 1)} \sum_{k>j} \frac{1}{S_{ijk}} a_{ij} a_{ik}$$

where S_{ijk} is the size of the shortest cycle which passes through vertices i , j and k . Note that if vertices j and k are connected, the smallest cycle is a triangle and $S_{ijk}=1/4$. If there is no loop passing through i , j , and k , then these vertices are treelike connected and $S_{ijk} \geq 1$. The cyclic coefficient of a network is the average of the cyclic coefficient of all its vertices:

$$\theta = \frac{1}{N} \sum_i \theta_i$$

In Recommender System, for example, we consider a network of scientific research collaborations in the area of Computer Science that has been published in the time frame of let's say 2017 to 2019. The nodes made to represent the Computer scientists. The computer scientist is connected if they have authored a joint paper or have a common sub research interest area. The cyclic coefficient will tell which vertexes have high connections with each other. Thus by measuring the distribution of the local cyclic coefficient, we can understand the details of the cyclic structure in the complex networks.

2.3 Transitivity

Transitivity, which is also known as Clustering, is a typical property of acquaintance networks, where two (2) individuals with a familiar friend are likely to know each other [1]. In the perspective of Recommender Systems customers that have some common items A_1 , A_2 are likely to purchase A_3 if bought by any of the other friends. Regarding a generic graph G , transitivity means the presence of a high number of triangles. This can be quantified by defining the transitivity T of the graph as the relative number of transitive triples, i.e., the fraction of connected triples of nodes (triads) which also form triangles.

$$T = \frac{3 * \# \text{ of triangles in } G}{\# \text{ of connected triples of vertices in } G}$$

2.4 Bipartite Graph

Two different nodes are identified about nodes on a Recommendation System. Node A may represent an individual, and node B may represent an interest (i.e., items). In a graphical form, as shown in fig. 2 below the Bipartite Graph is depicted as two(2) parallel planes where the nodes are interconnected by some links[2]. Each link connecting a node to another is a representation of common interest to both nodes. For example, both Node A and B bought an item the same mobile phone (iPhone, Samsung, etc.) and by this past data, the two(2) nodes are connected. The graph in itself is not typically usable by the learning algorithm. However, this helps to establish the relevant relationship between nodes knowing the common data they share.

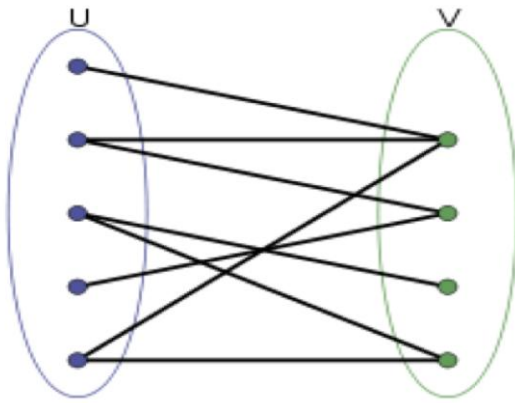


Fig 2: A Bipartite Graph

2.5 Centrality

Centrality measures are utilized to examine the complex network and to gauge the significance of a vertex in the system. Numerous centrality measures are proposed to assess the significance of a node or an edge in a complex network. In network analysis, three major types of centrality measures are used; namely Degree centrality, Closeness, and Betweenness.

2.5.1 Degree Centrality

The Degree centrality measure checks what number of neighbors a vertex has[12]. If the system is directed, two variants of the measure are considered. That is, the in-degree and the out-degree. The in-degree is the total count of the number of in-coming connections or the total number of successor vertexes. The out-degree is dictated by the vertex or node itself.

Let $A = (a_{i,j})$ be the adjacency matrix of a directed graph. The in-degree centrality x_{i} of node i is given by: $x_{i} = \sum_k a_{k,i}$ or in matrix form ($\mathbf{1}$ is a vector with all components equal to unity): $\mathbf{x} = \mathbf{1} A$ The out-degree centrality y_{i} of node i is given by: $y_{i} = \sum_k a_{i,k}$ or in matrix form: $\mathbf{y} = A \mathbf{1}$.

2.5.2 Closeness Centrality

Closeness Centrality of a node in complex network is determined as the inverse of the addition of the shortest paths between the vertex and all of the other vertexes in the network. Accordingly, the more focal a node is, the closer it is to every single other node.

$$C(x) = \frac{1}{\sum_y d(y,x)}$$

where $d(y,x)$ is the distance between vertices x and y .

The normalized form of closeness centrality is more often preferred, which is a representation of the average length of the shortest paths on the graph rather than the sum of the paths. The normalized is given as

$$C(x) = \frac{N}{\sum_y d(y,x)}$$

where N is the total count of the number of nodes in the graph.

2.5.3 Betweenness Centrality

The Betweenness Centrality(BC) property determines the noticeable quality of vertex or node in the graph. The BC shows the total count of the number of shortest paths that goes through a node. A node has a high Betweenness Centrality if

it lies on a high count of shortest paths between any two vertexes. The hub with high betweenness ordinarily has more prominent command over correspondence. They are additionally the ones whose expulsion from the graph will most disturb interchanges across different vertexes since they serve as the major hub for several messages.

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass-through v

3. CONSTRUCTING A RECOMMENDER SYSTEM

3.1 Techniques Used in Recommender Systems

There are several techniques (Algorithms) used to implement Recommendation Systems. The methods are Content-based Filtering Systems, Collaborative Filtering Systems, Demographic Filtering Systems, Knowledge RS, Utility RS, and Hybrid Recommender Systems [7][10][11]. Several works have discussed these techniques and stating some limitations that they have and the solution.

3.1.1 Collaborative Filtering

To make recommendations using the Collaborative filtering approach, the recommendations are made based on a few customers who are most similar to the active users [4]. It measures the resemblance of two users in various ways; one usual technique is to calculate the cosine of the space(angle) between the two vectors.

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \cdot \|\vec{B}\|}$$

The collaborative filtering technique relies on a social concept where users who have items presently will have yet common items in the future. For example, if user A, B, and C have both bought iPhone 3gs, 6s, and 8max in the past. Then when A, B purchases iPhone Xmax and C has not purchased, the recommender system will suggest iPhone Xmax to C. This because they have had a common taste on phones in the past and it may be true for the future also.

3.1.2 Demographic Filtering

The Demographic Filtering technique feeds on demographic information of users to make recommendations[6]. For example, John(male, age=14) bought some iPhone Xs and sneakers from Amazon, Kingsley(female, age: 16) bought items iPhone Xs and sunglasses from the same Amazon website, and Arthur(male, age=14) bought iPhone Xs. In this scenario John and Arthur have a common demographics of male and gender, there the system will recommend sneakers as an item to Arthur in his next purchase.

3.1.3 Context-Aware

The Context-Aware Recommender System feeds on information such as the location of the user, Identity of people the around, the date, the season, the temperature, and other related data to make recommendations. This technique is seen to be more precise and accurate[5]. For example, watching a movie: time, location, companion, emotions, occasion. Upon this contextual data, movies will be recommended to the active user based on the context data he/she fits into.

3.1.4 Content-Based

In the perspective of Content-based recommender systems, the items are represented by its related information or features[4]. The system also learns to know what the user has liked, or rated or ranked and uses that to represent the interest of the user. This system relies on keywords that are used to describe or identify the items. Therefore, the underlining algorithms will recommend items to a current user based on the best-matching of the previous rated items with similar existing ones[10].

3.1.5 Utility

The Utility based recommender algorithms make suggestions of items dependent on calculation of the utility of each item for the current user. The focal issue for this kind of framework is how to make a utility for individual users. In this recommender system, each industry will have an alternate system for touching base at a client explicit utility capacity and applying it to the candidate items. The primary preferred position of utilizing a utility-based recommender framework is that it can factor non-item qualities, for example, unwavering seller quality and item accessibility, into the utility calculation. This makes it conceivable to check the continuous stock of the item and show it to the client.

3.1.6 Knowledge-Based

The knowledge recommender system is one of the most powerful techniques in suggesting items to the user. The algorithms rely on the information of an item and relate it to the need of the user and by that makes a recommendation. The Knowledge RS gets information about the user's preferences[10][11]. For example, a book recommender system, the system may want to know why the user is buying a programming book. Is the user's level of programming imperative or the user just wants any write up approach using Objects Oriented Principle? By this given information, the recommender system establishes a relationship between the need of the user and the item to suggest books to the targeted user.

3.1.7 Hybrid

The Hybrid Recommender system utilizes a mix of the previously mentioned strategies by adopting the benefits of one procedure to remunerate the weaknesses of another, along these lines improving the general execution. Hybridization might be actualized in a few different ways: For example, by making content-based and collaborative filtering techniques recommendations independently and afterward combining them; or by adding content-based technique to a collaborative based methodology; or by bringing together the methodologies into one model.

4. ALGORITHM MEASURING TECHNIQUES

Knowing the similarity between item-item, item-customer, customer-customer is very crucial in Recommender Systems. Majority of the types relies on the ability of the recommender's algorithm to do this distinction. Common Neighbors and SimRank are both measuring techniques used in Complex Networks to measure the similarity of nodes and closeness of nodes which will be relevant in the make of Recommender Systems.

4.1 Common Neighbors

For a node x , let $\Gamma(x)$ denote the set of neighbors of x . In practice, two(2) nodes, x , and y , are more likely to have a link

if they have many common neighbors. The simplest measure of this neighborhood overlap is the directed count, namely

$$S_{xy}^{CN} = r(x) \cap r(y)$$

where $|Q|$ is the cardinality of the set Q . It is obvious that $s_{xy} = (A^2)_{xy}$, where A is the adjacency matrix: $A_{xy} = 1$ if x and y are directly connected and $A_{xy} = 0$ otherwise. Note that, $(A^2)_{xy}$ is also the number of different paths with length 2 connecting x and y .

4.2 SimRank

SimRank is defined in a self-consistent way, according to the assumption that two nodes are similar if they are connected to similar nodes.

$$S_{xy}^{SimRank} = C \cdot \frac{\sum_{z \in r(x)} \sum_{z' \in r(y)} S_{zz}^{SimRank}}{k_x \cdot k_y}$$

where $S_{xx}=1$ and $C \in [0,1]$ is the decay factor, the SimRank can also be interpreted by the random walk process, that is, $S_{xy}^{SimRank}$ measures how soon two(2) random walkers, respectively starting from nodes x and y , are expected to meet at a certain node.

5. BUILDING RESULTS

In the above sections, some theories and principles applicable to Complex Networks. In this session, the attention is on how to measure the performance of the algorithm that is going to be implemented.

The score values are assigned to the recommendations; that which is closer to one(1) is more commendable. Thus, the more the target item is in the top part of the list, the more the score approximates to 1.

$$\text{Score} = 1 - \frac{\text{Pos}_{item}}{\#items}$$

6. CONCLUSION AND DIRECTION FOR FURTHER WORK

Complex Networks is a system made up of several nodes that interact in a defined structure. The structural attributes of Complex Networks are highly identical with the underling techniques employed in Recommender Systems. Complex Networks have structural attributes such as the Small World Effect, Cyclic Coefficient, transitivity, Bipartite Graph, and Centrality. The algorithms of Recommender Systems such as Content-based Filtering, Collaborative Filtering, Demographic Filtering, Utility, Knowledge-based, and Hybrid methods relies on the concept of an item-item or item-customer relations. The item-item relation or item-customer relationships forms a logical network of nodes just as that of the Complex Networks. Upon this, it is argued that, the principles of Complex Networks are application in the construction of Recommender Systems.

In this research work, some number of techniques such as Common Neighbors, and SimRank were considered, which are relevant to the measure of the adopted methods of the Complex Networks in the recommendation process. Amongst the list of potential recommender items, a score is assigned. Items with assigned values that are closer to one(1) are top ranked. The more closer the value is to one(1), the more a better option it is for a recommendation.

To improve this research work, the discussed Complex Network attributes will be implemented to measure against baselines methods of Recommender Systems to measure percentage of improvement. In subsequent works, the various

7. REFERENCES

- [1] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D. U. “Complex Networks: Structure and Dynamics”, Physics Reports, Volume 424, Issues 4–5, February 2006, Pages 175-308
- [2] Zanin, M., Cano, P., Buldú, J. M., Celma, O. “Complex Networks in Recommendation Systems”, in International Conference on Computer Engineering And Applications (CEA'08) Acapulco, Mexico, January 2008pg 25-27
- [3] Shah, L., Gaudani, H., Balani, P. “Survey on Recommendation System” in International Journal of Computer Applications, Vol. 137, No. 7, March 2016
- [4] Nagarnaik, P., Thomas, A. “Survey on Recommendation System Methods”, in IEEE Sponsored 2nd International Conference On Electronics And Communication System, 2015
- [5] Schering, A., Düffer, M., Finger, A., Bruder, I. “A Mobile Tourist Assistance and Recommendation System Based on Complex Networks” in CNIKM, ACM, November 2009.
- [6] Costa, L. D., Rodrigues, F. A., Traverso, G., and Boas, P. R. V. “Characterization of complex networks: A survey of measurements”, in Advances in Physics, Vol. 56, No. 1, February 2007, 167–242
- [7] Jain, S., Grover, A., Thakur, P. S., Choudhary, S. K. “Trends, Problems And Solutions of Recommender System”, in International Conference on Computing, Communication and Automation (ICCCA2015)
- [8] Kearns, M. and Rayfield, R., Squash Magazine: The Small World of Squash. 2015[online]. Available at <http://squashmagazine.ussquash.com/2015/10/the-small-world-network-of-squash/>. [Accessed: 7-Jan-19]
- [9] Geeksforgeeks: Check whether a given graph is Bipartite or not. 2019[online]. Available at <https://www.geeksforgeeks.org/bipartite-graph/>
- [10] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. Mobile recommender systems in tourism. Journal of Networks and Computer Applications. 9, (2014) 319–333 pages. <http://dx.doi.org/10.1016/j.jnca.2013.04.006>
- [11] P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.): The Adaptive Web, LNCS 4321, pp. 377 – 408, 2007. Springer-Verlag Berlin Heidelberg 2007
- [12] Vairachilai, S., Kavitha, M. K., and Raja, M. 2017. Analysis of Statistical and Structural Properties of Complex networks with Random Networks. Applied Mathematics and Information Sciences **11**, No. 1, pages 137-146