Entropy Based Sensitivity Analysis and Visualization of Social Networks

Ekrem Serin¹, Serdar Adali¹, and Selim Balcisoy¹
Sabanci University, Istanbul, Turkey

Abstract. This paper introduces a technique to analyze and visualize a social network using Shannon’s entropy model. We used degree entropy and presented novel measures such as, betweenness and closeness entropies to conduct network sensitivity analysis by means of evaluating the change of graph entropy via those measures. We integrated the result of our analyses into a visualization application where the social network is presented using node-link diagram. The size of visual representation of an actor depends on the amount of change in system entropy caused by the actor and color information is extracted from the graph clustering analysis. Filtering of edges and nodes is also provided to enable and improve the perception of complex graphs. The main contribution is that the information communicated from a social network is enhanced by means of sensitivity analyses and visualization techniques provided with this work.

1 Introduction

Social network analysis has been actively studied in recent years [11], [3], [5], and has applications in many areas including organizational studies, social psychology and information science. The goal is to distinguish and detect regular or non-regular patterns, tendencies, mutual interests and reveal hidden information to execute the required tasks by perceiving the information presented.

The analytical social network analysis depends on information and a measure to quantify it. Here we borrow the concept of entropy which is introduced by Shannon[1] and apply it to measure the sensitivity of the social network. Entropy is a model for a general communication system and shows the uncertainty as well as the information amount of a system. In this work we consider entropy as information quantity and treat the social network as a system of communication. We calculate the system changes for each actor by removing that actor from the network and recording his/her change caused after removal. Hussain et.al.[3] treat entropy as uncertainty and apply it to calculate Bayesian posterior probabilities for discovering key players.

The social network used in this work is a scientific collaboration network extracted from DBLP[2] database including submissions for IEEE Transactions on Visualization and Computer Graphics(TVCG) between 2005-2009. We conducted sensitivity analysis for the collaboration network using degree, betweenness and closeness entropies. In order to present the aggregate entropy change,
each centrality measure entropy vector is normalized before combination process. Key actor discovery [3] is also integrated into the application.

In this work a visualization system that shows the collaboration network as a set of ellipses and arcs is provided. The size of the node can depend on any of the centrality, centrality entropy or aggregate entropy change measures. The color information is extracted from the result of graph clustering analysis.

The aim of a visualization is conveying some information to one that looks at it. We tried to develop a presentation system to show useful and meaningful information to the user with this work. Here we exploited Shannon’s entropy model as information amount inferred from a system, hence it is a social network in our approach, and tried to improve the user perception by conducting sensitivity analyses and providing a visualization application with different filtering mechanisms. We believe that the techniques provided here can help users to understand large-scale social networks or graphs by presenting useful information.

Our contributions are, a novel approach for the sensitivity analysis of a social network and a visualization system that demonstrates the information deduced from the sensitivity, key player and clustering analyses.

The rest of the paper is organized as follows, in Section 2 we discuss about the related work, in Section 3 we present entropy based sensitivity analysis of social network, in Section 4 we discuss about the visualization system and deliberate on the outputs, and Section 5 concludes our work.

2 Related Work

In recent years many methods have been developed for social network analysis to rank nodes, to discover hidden links, to deduce meaningful information by the help of statistical, dynamic or visual perspective analyses[4]. The context of social network analysis varies from dark networks[5], to collaboration networks [6] or to networks in biological sciences.

Statistical analysis of social networks uses statistical properties of graphs including clustering, degree distributions or centrality measures to deduce useful information. Centrality measures determine the relative importance of a node in a network and the most common ones are degree, betweenness and closeness[8]. A more complex measure i.e. Markov centrality[9] treats the social network as a Markov chain and helps to discover significant facilitators in that network.

Choosing the right centrality for a specific problem is usually a hard task and common approach is comparing different centralities for the same network and building hypothesis about the discovered central nodes [10].

One of the pioneers in exploring key actors for dark networks Sparrow[7] used six centrality measures for their relevance in revealing the mechanics and vulnerabilities of criminal enterprises. Hussain et.al.[3] used degree centrality measure to set Bayesian Posterior Probabilities for entropy change calculations to locate key actors in social networks. Newman[6] defined a different set of statistical measures such as number of authors, mean papers per author, mean
authors per paper, number of collaborators, and average degrees of separation for scientific collaboration networks. Crnovrsanin et.al.[11] used Markov centrality metric to discover and highlight meaningful links.

Another aspect of social network analysis is to discover the dynamic behaviors of the network which usually takes domain of time into account. Dynamic analysis can include network recovery by multiple representations from longitudinal data to model the evolving network, network measurement of deterministic, probabilistic and temporal aspects and statistical analysis such as continuous Markov model, and Cox regression analysis for determining significant nodes.

Kaza et.al .[5] used multivariate survival analysis of Cox regression for significant facilitator discovery. Falkowski et.al.[12] proposed a technique to detect the evolution of subgroups and analyzing subgroup dynamics in manner of stability, density, cohesion and distance using temporal and statistical analyses.

3 Sensitivity Analysis

The sensitivity of an actor in the social network reveals the importance of relation between the actor and all other participants. Here we present an analytical approach using centrality entropy distributions which can be considered as good indicators of network sensitivity. We define three centrality entropy distributions, degree entropy, betweenness entropy and closeness entropy. Mutual information is presented by the normalization of centrality entropy distributions exercised in this work. Subsections will deliberate on the centrality entropies via the help of Shannon Entropy.

3.1 Degree Entropy

The Shannon entropy [1] of a discrete random variable $X$ with values in the set \{a_1, a_2, ..., a_n\} is defined as

$$H(x) = \sum_{i=1}^{n} p(x_i) I(x_i) = - \sum_{i=1}^{n} p(x_i) \log_b p(x_i)$$  \hspace{1cm} (1)

In equation (1) $p(x_i)$ is the probability mass functions of state $x_i$, for a system with $n$ different states. In our context the probability mass function set is the degree distribution of the actors in the social network and $n$ is the number of distinct actors. Hence each edge connects two nodes, that edge is counted for both actors. The probability mass function $p(x_i)$ of the node $x_i$ is defined as,

$$p(x_i) = \frac{E(x_i)}{\sum_{j=1}^{n} E(x_j)}$$  \hspace{1cm} (2)

In equation(2) $E$ is the number of edges in the graph. In order to conduct sensitivity analysis using degree entropy the initial information amount, hence degree entropy is recorded including all the actors in the social network. An actor
is removed from the network and the system entropy is re-calculated for the remaining actors. To calculate the system entropy we use the largest connected component of the subgraphs if the actor disconnects the network. The calculated entropy value is recorded and actor is connected to the network. This sequence is applied to all actors in the social network. A change analysis for each actor is performed by taking difference of initial system entropy and remaining system entropy. The result is normalized and sorted from the most to the least.

### 3.2 Betweenness Entropy

The betweenness centrality measure is defined as the number of shortest paths from all vertices to all others that pass through that node, hence it is

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

Since Shannon Entropy model is mentioned in previous section, only the probability mass function for betweenness entropy is deliberated here, which is the betweenness centrality distribution of the actors scaled for undirected networks shown in equation(2).

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \frac{(n-1)(n-2)}{2}$$

The sensitivity analysis using betweenness entropy is done similar to the degree entropy analysis. The initial system entropy for betweenness probability mass functions is calculated and recorded, and each actor is removed from the network where the change presented by that actor is recorded. After the recordings, the differences are calculated and normalized.

### 3.3 Closeness Entropy

The closeness centrality of a node measures how easily other nodes can be reached from it or how easily it can be reached from the other nodes. It is defined as the number of nodes minus one divided by the sum of the lengths of all geodesics from/to the given node shown in equation(5).

$$g(v) = \frac{n - 1}{\sum_{t \in V \setminus v} d_G(v,t)}$$

We used the values calculated in equation(5) as the probability mass function for the equation(1) to compute closeness entropy for the social network. The sensitivity analysis is done using the sequence presented in previous sections however in this case closeness entropy is used as probability mass function.
3.4 Combined Approach

Degree, betweenness and closeness entropies are combined to measure the aggregate sensitivity of each actor in the network. The combination is product of the normalized changes of the centrality entropy values. This value incorporates the information about three centrality measures in a single data.

\[ I(v) = I_d(v) \cdot I_b(v) \cdot I_c(v) \] (6)

In equation (6), \( I_d(v) \) denotes degree change information, \( I_b(v) \) denotes betweenness change information and \( I_c(v) \) is closeness change information where we treat information as the system entropy. The user can select any of them as well as the combined one for further analysis using the visualization system provided with this work.

4 Visualization and Discussion

There are many techniques found in literature [13] for social network visualization varying from node-link diagrams, to tree-maps, from adjacency matrix representations[14] to sophisticated 3D visualizations, however we believe that node-link diagrams are most suitable presentation of social networks for human perception.

In this work, we provide a visualization application that presents social network as node-link diagram. Centrality measures, centrality measure entropy changes i.e. sensitivities are conveyed to the user via drawn nodes. For instance if an actor changes the system entropy more than the other actors, that actor is represented with a greater ellipse. We also integrated the key actor locating algorithm to our visualization software. The layout and clustering analysis is done using the energy-based minimization model presented by Noack [15].

TVCG(2005-2009) collaboration network is visualized in Figure 1 using different information and filtering applied to nodes and edges. Figure 1.(a) shows the default presentation of the network, no information except the connectivity of the actors is conveyed to the user. In Figure 1.(b) clustering information is applied and edges are filtered out, the user can percept the groups from the presented colorized picture. Figure 1.(c) presents the result of sensitivity analysis of degree entropy applied to the nodes sizes with clustering colorization, the user can deduce each actor’s effect using degree entropy to the whole system, hence the actors shown with a greater ellipse changes the system entropy more than the other actors due to the degree centrality. In Figure 1.(d) sensitivity analysis of combined centrality entropies is shown, here we combined degree, betweenness and closeness entropies using the technique discussed in previous section. The user can percept each actors change to the system using the combined information, hence some actors represented with smaller or bigger ellipses depending on the total change they caused. In Figure 1.(e) the sensitivity analysis of degree entropy is applied with node filtering using visual transparency. Here the
demonstration of transparency filtering is exercised, however the user can select to completely undraw the the actors below some filtering threshold that they caused to the system change. In Figure(f) the result of key actor location algorithm[3] is presented with colorized clustering information.

The sensitivity analyses conducted using centrality measure entropies show the changes to the system entropy caused by the actors in the network. The cause of change differs by the amount of information decreased from the initial information calculated for the system depending on the used centrality measure entropy. The change factor depends on two criteria, the first one is the number of disconnected nodes caused by the actor after removal, the second one is the centrality measure entropy amount of the disconnected actors, which actually complies with the aim of sensitivity analysis that is revealing importance of relation between the actor and all other participants in the system.

5 Conclusion

In this paper a technique for analyzing and visualizing a social network using Shannon’s entropy definition is presented. We used the three most common centrality measures such as degree, betweenness and closeness to define centrality measure entropies. Centrality measure entropies are utilized to conduct the sensitivity analysis of system employing entropy changes of the actors in the social network.

We tried to enhance the information communicated from a social network by help of analyses and visualization techniques provided in this work. Experiments are preformed using different datasets varying from hand generated to collaboration data extracted from various sources. A social network example TVCG collaboration data is presented here to show the results of our work.

Our experiments have shown that Shannon’s entropy model is a promising way to analyze and visualize social networks by providing a measure to quantify the information on the communication channel between the user and visual world in computer.

References

2. The DBLP Computer Science Bibliography, http://dblp.uni-trier.de/
Fig. 1: TVCG(2005-2009) collaboration network is visualized using different information and filtering applied to nodes and edges. (a) shows the default presentation of the network. In (b) clustering information is applied and edges are filtered. (c) presents the result of sensitivity analysis of degree entropy applied to the nodes sizes. In (d) sensitivity analysis of combined centrality entropies is shown. In (e) the sensitivity analysis of degree entropy is applied with node filtering using transparency. (f) presents the result of key actor location algorithm.