Computing the value of a crowd

Manas S. Hardas[1, 2] and Lisa Purvis[1]

mhardas, lpurvis@spigit.com

Spigit, Inc. 311 Ray Street, Pleasanton, CA 94566[1]
Computer Science Department, Kent State University, Kent, OH 44240[2]

Abstract. This paper proposes that the "value" of a crowd can be defined in terms of the overall engagement of the individuals within the crowd and that engagement is a function of certain characteristics of the crowds such as small world-ness, sparsity and connectedness. Engagement is hypothesized as messages being exchanged over the complex network which represents the crowd and the "value" is calculated from the entropy of message probability distributions. An initial random network is passed through a process of entropy maximization and the values of some structural properties are recorded with the changing topology to study the corresponding behavior. We show that as the small world-ness and connectedness of a crowd increases and the sparsity decreases, the engagement in the crowd increases.

1 Introduction

In the recent past it is increasingly being observed that social media plays an important role in putting a dormant crowd in action. The most recent examples of social media fueled crowd activism are the Egyptian revolution, the Spanish May 15th movement [5] and the most recent Occupy wall street protests [2] spread out across many cities in the United States. The common underlying theme in all these examples of activism is the intelligent use of social media like Facebook, Twitter, etc. to bring a crowd into action. Beneath any crowd is a network of people which have to be engaged towards a common goal for a crowd to perform a purposeful action. The engagement of a crowd can be measured in terms of the activity between the individual nodes. If the "activity" between the nodes is high, it means that the crowd is better engaged. The "activity" between any two nodes can then be measured as the number of messages exchanged between the two nodes.

We propose that the value of a crowd comes from its engagement and a very well engaged crowd has more value than a crowd which is not engaged. Engagement has certain properties like working towards a common goal, working in parallel on different tasks to achieve a common higher task, etc. The value of a crowd changes from time to time. Some movements start off great but die out soon, some movements start out small but pick up momentum and turn into a full-fledged revolutions. Of course there are many reasons for this happening and in this paper we briefly try to focus on the structural aspects of the crowd network.

Hardas, M. and Purvis, L. Computing the value of a crowd, p. 8, 2011.

Crowd networks are very different from typical static and structured communication networks. Although underlying the crowd is a social network of connected individuals, the phenomenon of a crowd happens only in the presence of certain characteristic conditions. One of the central conditions is the messaging activity which happens between the individuals which form the crowd. The so-called value of a crowd comes from the ability of a crowd to interact with each other and communicate with each other by sending and receiving messages – in essence the engagement of the crowd with one another. It is because of this activity on the links of crowd networks that creates the power and capability to achieve valuable actions and results.

Crowds can be represented as complex networks with different topologies and internal structural properties like small worlds, community organization, component structure, sparseness, etc. Therefore crowds can also be analyzed for their complex network properties like the clustering in a network, shortest path lengths, size of the giant component, etc. This paper presents a hypothesis that the engagement (and the resultant value generated from the engagement) of a crowd can be measured in terms of the entropy of the network. The entropy is calculated as a function of the probability distributions of the incoming and outgoing messages, which represents the entropy or uncertainty in the activity over the network. In this way we can translate the activity occurring over a network in terms of message exchange as a measure of the value of a network. An evolutionary algorithm is presented to optimize the entropy of a network by successively changing the network topology. Results indicate that the value of the crowd is very closely related to its small world-ness, sparsity and connectedness.

2 The significance of engagement in Crowd Networks

One of the principles articulating the social value of a network is the Metcalfe's law ($value \propto n^2$) which states "the value of a communications network is proportional to the square of the number of connected nodes which grows geometrically as the number of nodes grow." In other words, as a network gets bigger it increases in value than it would by just adding more nodes. For instance, a network of 5 nodes would have a value of 25 while a 10-node network would have a value of 100 (double the number of nodes, but 4 times the value).

While Metcalfe's law provides a useful view of a network's value, it focused on more static, communication networks as its basis and makes a variety of assumptions on the type of structures in the network. Most real complex networks are not homogenously linked by similar type of edges. Most real world complex networks display certain characteristic properties like small world phenomenon and scale free distributions [1, 7]. Today's hyperconnected networks of people, information, and devices pose an entirely different challenge – in order to extract the value of today's crowd networks, it is not enough to understand simply the connectedness of the network. We must also understand the activity over those connections, representing the engagement of the crowd, and how it grows or shrinks over time and under what conditions. Only then can we begin to extract the true value of the crowd network. A crowd network has structural and behavioral complexities not present in a more static communication network. As such, crowd networks have significantly different dynamics and models of information and opinion spreading than a more predictable device communication network. Key differences are the social clustering that happens in a crowd network – effectively creating clusters of large strongly connected components, as well as a typically small amount of weak links that connect the entire network into one. Furthermore, the activity of communications across the connecting links as well as the triggers for growth or decline of that activity creates emergent network effects whose drivers today are largely unknown and undefined. Recent examples of real-life crowd network effects taking place are the May 15^{th} movement in Spain, the self-organizing response network after the World Trade Center attack on September 11^{th} 2001, and the Haiti earthquake response network. In each of these cases, we saw network effects at work, where the crowd drove results and action from the broad spreading of information, from a decentralized coordination and synchronization, and from the rapid information dissemination triggered in these instances. The value of the crowd response in each of these instances hinged on communication activity throughout the network, i.e. engagement and growth of that engagement. We next examine what characteristics of the crowd network structure could lead to this value and engagement over the network connections.

3 The value of a crowd

A crowd is defined as a complex network of people which can be represented by a graph consisting of vertices/nodes and edges/links. A vertex represents a person and an edge represents the connectivity between two people.

Let the crowd 'C' be represented by a graph $G = (V, E)$ where $V = \{v_1, v_2, \ldots, v_n\}$ and $E = \{e_1, e_2, \ldots e_m\}$ where n is the number of nodes and $m \leq \frac{n(n-1)}{2}$ is the number of edges. The graph is represented by an adjacency matrix A in which each element a_{ij} = [0, 1] denotes the existence of a link between nodes i and j. If no link exists then $a_{ij} = 0$. Each node is characterized by two probabilities. The first is the probability of a node to receive a message from any other node in the network and the second is the probability that the node sends a message to any other node. These probabilities are represented in a matrix of probability distributions in which each vector is associated with each node. The incoming probability distribution is represented by random variable $X = \{x_{ij} \mid i, j \leq n\}$ and the outgoing probability distribution is represented by a random variable $Y = \{y_{ij} \mid i, j \leq n\}$. The probabilities are stored in matrices $X_{n\times n}$ and $Y_{n\times n}$ respectively. Thus, the incoming probability of node v_i to receive a message from v_j is given by $x_{ij} \in X$ and the outgoing probability of node v_i to send a message to node v_j is given by $y_{ij} \in Y$.

3.1 Entropy of a network

The entropy of a network is the measure of the uncertainty in a network. It measures the network's heterogeneity in terms of the diversity in the incoming and outgoing message distributions. The value of a node is calculated in terms of the incoming and outgoing message entropy, which is the entropy of the incoming and outgoing probability distributions associated with a particular node. Note that the entropy here does not try to measure the actual information content of the messages being exchanged but only the randomness in the incoming and outgoing message probability distributions. The cumulative incoming and outgoing message entropies of a network are calculated as the summation of all the individual incoming and outgoing node entropies. Thus,

$$H^{in} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} * x_{ij} \log\left(\frac{1}{x_{ij}}\right) \quad \ldots(1)$$

$$H^{out} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} * y_{ij} \log\left(\frac{1}{y_{ij}}\right) \quad \ldots(2)$$

Finally, the total value of a network (crowd C) is calculated as a weighted measure of the incoming and outgoing entropies of the network. Thus, the value V_C is represented as a function of weighing variable α as in Eq.3. The variable α allows for weighing in the incoming and the outgoing network entropies.

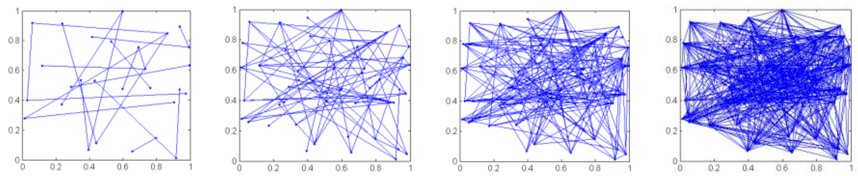
$$V_C(\alpha) = \alpha H^{in} + (1-\alpha)H^{out} \quad \ldots(3)$$

3.2 Entropy maximization algorithm

The main objective of the algorithm is to find a network topology which maximizes the entropy of the network (value of the crowd) in terms of the incoming and outgoing messages calculated as above. We start with a randomly generated adjacency matrix with a random number of connections according to a predefined probability p. The algorithm is as follows,

1. Calculate the value of the objective function (Eq. 3) for the network represented by the adjacency matrix A.
2. If the value is greater than the previous value of network entropy then accept the new network. Otherwise randomly flip each bit in the current adjacency matrix with a probability p to obtain a new adjacency matrix representing a new network topology.
3. Repeat step 1 and 2 until the entropy of the newly generated network does not exceed the previous value for at least n^2 times.

The limit of n^2 times to check for the best candidate is selected arbitrarily. Consequently an upper limit can be set for the acceptable number of links in a network as the stopping criteria. In experience, if allowed to proceed then the algorithm keeps on looking for better network topologies in terms of the given incoming and outgoing distributions. The probability distributions are generated uniformly at random. In theory, the maximum entropy is achieved when all the nodes in the network are. However, a fully connected network generated by this process has no meaning from the point of view of the structural regularities which we are trying to analyze. Therefore a threshold is set to stop the optimization process after a certain number of time steps. Other more sophisticated measures of the network entropy can be found here [3, 6, 8, 9].



Cumulative entropy →

Fig. 1. Depicts the evolution of the graphical structure through the process of entropy maximization. As the cumulative entropy increases, so do the number of connections.

3.3 Analysis of the structural properties

In the above given model, the entropy of a network is a function of the topological features of the network. The algorithm tries to maximize the network entropy such that more nodes communicate with each other [Fig. 1]. We observe the behavior of four parameters with respect to the cumulative entropy of the network. Fig. 2 shows the behavior of the structural parameter versus the network entropy.

Total number of active links (m). As the cumulative entropy increases so do the number of links in a network. This means that the value of a crowd increases with the number of people communicating with each other which seems to be self-evident.

Clustering coefficient (C). Clustering coefficient is defined as the fraction of nodes which form triangles in a network [7]. The formation of a triangle is indicative of how closely related neighbors of neighbors are inside a network. Again it is self-evident that to the value of a crowd increases with the increase in the clustering in a network. As more nodes begin forming clusters (of basic size 3 in this measure), this results in high local clustering and therefore results in better engagement.

Average shortest path length (l). This parameter measures the average distance in terms of number of hops between any two vertices in a network. It is observed that as the cumulative entropy increases the average shortest path length decreases from over 2 to 1.5. The value of a network increases if more nodes are accessible in shorter paths from each other.

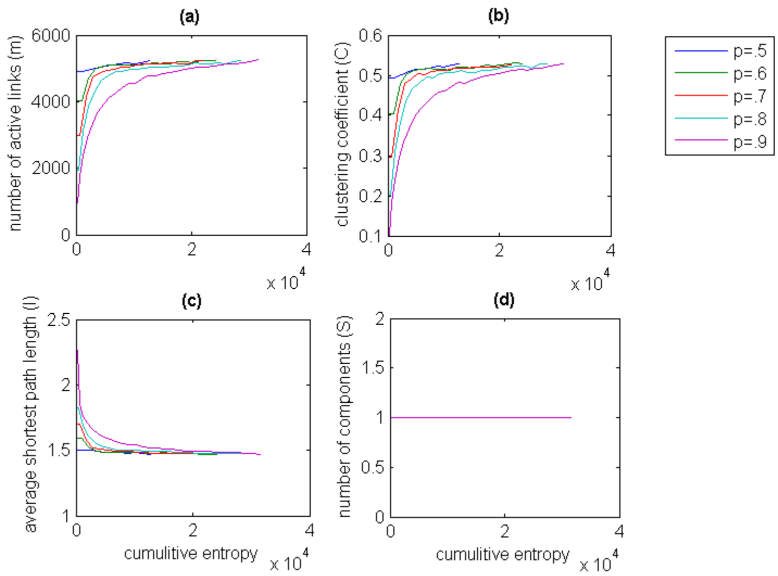


Fig. 2. Shows the behavior of structural parameters plotted against the increasing cumulative entropy of the network. For all the networks the number of nodes n=100 and the weighing parameter is constant at $\alpha = 0.5$. The link density probability (p) is varied from 0.5 to 0.9. For all values of $p < 0.5$, the parameters display more or less linear behavior. (a) The increase in the entropy is brought about by increase in the number of active links (m) in the network (b) shows as the number of links increases so does the clustering coefficient along with the network entropy, (c) shows the decrease in the average shortest path length along with the network entropy (d) shows number of disconnected components very quickly decreases to 1 forming a single giant component with increasing network entropy. In all cases after an initial sudden increase/decrease the values quickly normalize to the theoretical limits.

Number of components (S). Connected components are those nodes in a network which are accessible from each other directly or through a different path. A network may contain many components in which the nodes within a component are accessible to each other but one component is not connected to another. In this case the nodes in different components are assumed to be at infinite distance from each other. A network with a single giant component is obviously better from the point of view of crowd engagement since all the nodes in the graph are accessible through a path.

4 Discussion

Small world properties. Small worlds are characterized by low average short path lengths and high clustering. This phenomenon is observed when the average distance between two nodes grows logarithmically with the size of network. Small worlds have been observed in many real world networks. Watts and Strogatz [10] showed that small worlds existed in real world networks like the power grid of United States, collaboration of film actors, and the neural network of the worm C. elegans. Other real networks which show small world effect (reviewed in [10]) are, World Wide Web [1], collaborations in physics and mathematics literature [7], metabolic networks in biology [7], protein interaction networks [7], etc. From the above figures it is clearly seen that the entropy optimization process leads to networks which are small worlds with increasing clustering and decreasing shortest path lengths. Thus, the value of a crowd in term of network engagement is inherently associated with its small world-ness.

Sparsity. As the network gets denser in terms of the number of links the engagement increases. This means that the average number of nodes an individual node connects to increases with the entropy. This may seem self-evident as the number of links increase through the optimization process. The value of the crowd then is a function of the average number of nodes a node connects to (average degree).

Connectedness. The connectedness of a network is usually measured in terms of the number of components. As it is seen from the figures about, during the process of optimization the networks very quickly forms a single giant component consisting of all the nodes from a few disconnected components in the beginning as the entropy increases. This means that to achieve a higher value for a crowd, it is essential that most of the nodes be connected to each other through some path if not directly connected. Despite the sparsity of the network during the initial stages of optimization, it is observed the giant component is formed very quickly in proportion to the sparsity. Thus a connected crowd is a better engaged crowd.

5 Conclusion

It is hypothesized that certain statistical regularities seen in complex networks contribute to the "value" of a crowd as measured in terms of the messages being exchanged over the network. We show that indeed crowds which are better engaged do seem to be small worlds with high clustering and low average shortest path. They also generally seem to be densely connected into a single giant connected component. Understanding more deeply the aspects of crowd network structures that contribute to the overall collective value will enable new uses and applications of crowds and effective leverage of crowd behaviors for problem solving. In the future we would like to explore how certain influentials in the network contribute to the value of the crowd and how network phenomenon like percolation affect the value.

6 References

1. Barabási, A.L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286, 509-512.

2. Caren, Neal and Gaby, Sarah, Occupy Online: Facebook and the Spread of Occupy Wall Street (October 24, 2011). Available at SSRN: http://ssrn.com/abstract=1943168

3. Demetrius, L. "Robustness and network evolution--an entropic principle." Physica E, v. 346 issue 3-4, 2005, p. 682.

4. James Surowiecki. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, Brown publishing, 2004.

5. Javier Borge-Holthoefer, Alejandro Rivero, Iñigo García, Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study, PLOSOne August 19, 2011

6. K. Anand and G. Bianconi, Phys. Rev. E 80, 045102 (2009).

7. Newman, M.E.J. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98, 404-409.

8. Ramon Ferrer I Cancho and Ricard Sole. Optimization in Complex Networks, Lect. Notes Phys. 625, 114-125 (2003).

9. R. V. Sole and S. Valverde, Information Theory of Complex Networks: On Evolution and Architectural Constraints, Lect. Notes Phys. 650, 189-207 (2004).

10. Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature, 393, 440-442.