Entropic Centrality for non-atomic Flow Networks

Frédérique Oggier
Division of Mathematical Sciences
Nanyang Technological University
Singapore
Email:frederique@ntu.edu.sg

Silivanxay Phetsouvanh and Anwitaman Datta
School of Computer Science and Engineering
Nanyang Technological University
Singapore
Email:SILIVANX002@e.ntu.edu.sg,anwitaman@ntu.edu.sg

***Abstract*—Given a graph, the notion of entropic centrality was introduced by Tutzauer to characterize vertices which are important in the sense that there is a high uncertainty about the destination of an atomic flow starting at them, assuming that at each hop, the flow is equally likely to continue to any unvisited vertex, or to be terminated there. We generalize this notion of entropic centrality to non-atomic flows, and furthermore show that the case of a non-atomic flow splitting with equal probability across different subsets of edges results in the same entropic centrality as that of the atomic flow. This gives a new and more generalized interpretation to the original entropic centrality notion. Finally, we demonstrate using network graphs derived from Bitcoin transactions that depending on the graph characteristics, the presented entropy based centrality metric can provide a unique perspective not captured by other existing centrality measures - particularly in identifying vertices with relatively low out-degrees which may nevertheless be connected to hub vertices, and thus can have high spread in the network.**

I. INTRODUCTION

Given a graph to be analyzed, vertex centrality is a notion meant to identify the most important vertices within this graph. What “most important” means is not universally defined, therefore, numerous notions of centrality have been proposed [1], e.g.: the *degree centrality* is just the degree (or in-degree/out-degree) of the vertex depending on whether the graph is directed, possibly normalized to get the fraction of vertices a given vertex is connected to; the *closeness centrality* is the reciprocal of the sum of the shortest path distances from a given vertex to all others, typically normalized; the *betweenness centrality* is the sum of the fraction of all pairs of shortest paths that pass through it.

The closeness centrality gives the highest weight to vertices which are closer to all others. The betweenness centrality characterizes the number of times a vertex acts as a bridge along the shortest path between two other vertices. It was introduced in social networks [2] to quantify the control of a human on the communication between other humans. The degree centrality can be interpreted in terms of immediate propagation of a flow from a given vertex. This is of interest if say we are considering the spread of an epidemic.

Centrality from a flow view point has been investigated in [3], where different types of flows were categorized into parallel duplication (e.g., email virus alert), serial duplication (e.g., person-to-person gossip), or transfer process (e.g. package delivery), and the methods by which the traffic spreads were identified to be paths, geodesic, trails, and walks.

The idea to look at flows from an information theoretic view point was proposed in [4]. The larger the uncertainty in determining the destination of the flow, the more central the node of origin is deemed. This gives rise to the notion of *entropic centrality*, which is the focus of this paper. In [5], the model of [4] was slightly modified so that the flow propagation can be modelled as a Markovian transfer process. Both [4] and [5] assume that the flow is atomic, that is, when it arrives at a graph vertex, it can either stay there, or can go to one of the neighbors, but cannot be split into different neighbors.

The contribution of this paper is to extend the framework of [4] to a case where the flow at a given vertex can split among a subset of adjacent vertices:

(1) We propose in Section II a flow model that allows the flow on a graph to split, based on which we generalize the entropic centrality notion defined in [4].

(2) We prove in Section III that (perhaps) surprisingly, a case of the entropic centrality that we propose for non-atomic flows, namely when the flow splits across subsets of neighbors uniformly at random, simplifies to the entropic centrality proposed in [4], which in turn gives a new and more powerful interpretation to this centrality measure.

(3) We provide examples and explore applicability and limitations of the entropic centrality measure in Section IV, elaborated using subsets of the Bitcoin transaction network.

II. THE SPLIT-AND-TRANSFER FLOW MODEL

Let $G = (V, E)$ be a graph which is directed and connected, with finite vertex set V. The set $E = \{(u, v),\ u, v \in V\}$ of edges describes a relation on V, which we assume is reflexive, that is, every vertex has a self-loop. The case where the relation is symmetric is subsumed, since it suffices to have (v, u) in E whenever (u, v) is. For a graph which is not connected, each connected component can be treated separately. Since G is directed, every vertex v has an out-degree $d_{out}(v)$, counting how many edges are “exiting” v, including the self-loop.

A. Network Flows

Let $d : V \to \mathbb{R}$ be a function such that $\sum_{v\in V} d(v) = 0$. It is called a *demand function*. Vertices for which $d(v) = 0$ are called transit vertices. By a *flow*, we mean a function $f : E \to \mathbb{R}^+$ with the property, called flow conservation, that $d(w) = \sum_{v,\ (w,v)\in E} f(w, v) - \sum_{v,\ (v,w)\in E} f(v, w)$ for all $w \in V$. In a classical flow network scenario, the demands

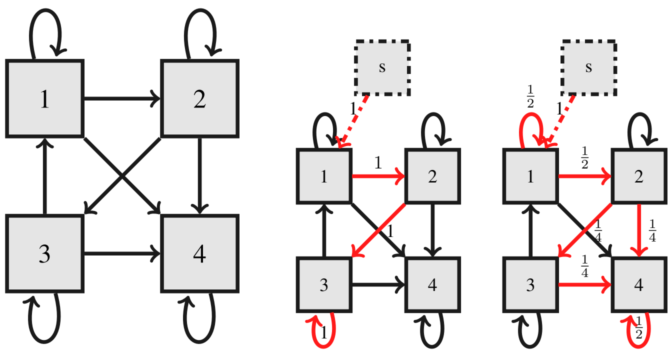


Fig. 1. A small directed connected graph $G = (V, E)$, with $V = \{1, 2, 3, 4\}$, $E = \{(1,1), (1,2), (1,4), (2,2), (2,3), (2,4), (3,3), (3,4), (4,4)\}$. The out-degrees are $d_{out}(1) = 3$, $d_{out}(2) = 3$, $d_{out}(3) = 2$, $d_{out}(4) = 1$. Two instances of flow propagation are given on the right (each instance will then happen with a given probability). Flow values are shown on the edges.

are assigned to every vertex once and for all. Here, while the underlying graph G remains the same, we will consider at the same time different instances of demands, and compute the probability that a flow achieves them. In all cases, there will be only a single vertex with positive demand, which can be considered as the source s of the flow.

Suppose that we want to study the centrality of a vertex $u \in V$. We introduce an auxiliary vertex s that serves as the source, and add an auxiliary edge that connects s to u. Then $d(s) > 0$ and the flow conservation tells us that $d(s) = \sum_{v,\ (s,v)\in E} f(s,v) = f(s,u)$ that is the amount $f(s,u)$ at s flows from s to u. We may normalize the original flow to be 1, i.e., $f(s,u) = 1$. The reason for this formalism is that we assume that when the flow f starts at some given vertex $u \in V$, it can remain at u (that is, use the self-loop), or partly remain at u and furthermore branch out to go to any subset of adjacent vertices of u (see the right flow instance on Fig. 1), or just either split-and-transfer and transfer to any subset of adjacent vertices of u (see the left flow instance in Fig. 1 for a transfer example). When a splitting happens, the amount of flow on each branch is a fraction of the incoming flow, such that the sum of the outgoing flows satisfies flow conservation. The first scenario means that $d(u) = -f(s,u) = -1$, the second that $d(u)$ is minus some fraction of $f(s,u)$ ($d(1) = -\frac{1}{2}$ in the right flow instance on Fig. 1, so that flow conservation $d(1) = -\frac{1}{2} = \frac{1}{2} - 1$ holds), and the third that u is a transit node ($d(1) = 0$ in the left flow instance on Fig. 1).

At the next iteration of this process, when the flow leaves these adjacent vertices of u, possibly branching out again, we assume that none of those branches are allowed to return to u, and generally, at any iteration, any branch will ever go back to a vertex previously visited in the path of the branch itself - thus to say - there will be no cycle formed by a branch. Consequently, upon reaching a vertex whose neighbors have all already been visited by the branch of the flow, that branch will necessarily stop, and the vertex will have a negative demand. For example, on the left flow instance in Fig. 1, the flow reaches vertex 3, and cannot go back to vertex 1 (it could have continued to node 4, but it does not). The above conditions do not prevent the overall flow to have cycles, and branches followed by the flow may intersect and/or merge, as is the case on the right flow instance in Fig. 1, where the branch $2 \to 3 \to 4$ merges with the branch $2 \to 4$.

Finally, a branch of the flow could also stop at any iteration, which is represented by a self-loop on the vertex reached by this branch at this iteration, absorbing the corresponding part of the flow. For example, on the left flow instance of Fig. 1, the branch $1 \to 2 \to 3$ stops at 3, and $d(3) = -1$.

This above definition of *split-and-transfer* flow is similar to that of flow in uncapacitated networks [6]. All vertices are transit vertices, except for the source s which supplies the flow, and the vertices where branches of the flow (partly) stop. The present setting allows to consider different flows, leaving different vertices, and reaching correspondingly different subsets of vertices, while keeping the same underlying graph, and the demand function captures flow volume conservation. The split-and-transfer model is suitable to describe networks induced by various social activities, such as distribution of goods that can be sold in different amounts, or monetary transactions, as we will later describe in the case of the Bitcoin network.

B. Split-and-Transfer and Stop Probabilities

Now to every vertex v, we attach a probability that describes how the flow that arrives at v will behave next. Let $d_{out}(v)$ be the out-degree of v, which includes v itself. In [4], an atomic flow was considered (so there was no branching of the flow at any stage), and a transfer probability was defined to be $\frac{1}{d_{out}(v)}$, which is the probability to choose one outgoing edge uniformly at random, possibly the vertex v itself. It was further assumed that the flow makes no cycle, thus if the specific flow path considered has already a_v of v's neighbors visited previously, the effective number of remaining choices is $d'_{out}(v) = d_{out}(v) - a_v$, and accordingly, the probability, given the chosen path, becomes $\frac{1}{d'_{out}(v)}$. The scenario where the flow stops at v is modelled by the self-loop, which is one choice among all the others. Naturally, the value of this probability of stopping is 1 when all the neighbors of v have already been visited by the flow in question.

Suppose now we want the flow to not only be transferred from one vertex to another, but to also be possibly split across different vertices, while conserving the total flow volume. Given a vertex v, let $\mathcal{P}_{s,v}$ be the set of paths from s to v, and let $\mathcal{P}_{P_v,w}$ be the set of paths from v to w, where w is any allowed end of paths given an upstream path $P_v \in \mathcal{P}_{s,v}$. Given such a path $P_v \in \mathcal{P}_{s,v}$, v has a set $\mathcal{N}(P_v)$ of allowed neighbors, and we can define a *split-and-transfer probability* $\tau_{P_v}(v)$ which is the probability that the flow goes out of v on a particular subset of $\mathcal{N}(P_v)$, given the path P_v was used to arrive at v. Then from v, the flow has $d'_{out,P_v}(v) = |\mathcal{N}(P_v)|+1$ effective number of choices to branch out (including self loop). The flow under consideration can then still continue to them from v. There are a priori $2^{d'_{out,P_v}(v)} - 1$ ways to choose a subset $\mathcal{S}(P_v)$ of $\mathcal{N}(P_v)$ for the branch of the flow to go to, this is the cardinality of the power set of $\mathcal{N}(P_v)$, without the empty set. If all the subsets are assigned equal probability, then $\tau_{P_v}(v) = \frac{1}{2^{d'_{out,P_v}(v)}-1}$. One of these choices is the self-loop indicating that the branch of the flow totally terminates at this vertex. In general, the stopping probability also has the same expression, and it becomes 1 at a vertex if all its neighbors have already been visited by the given branch of the flow. Since we want volume conservation during a split, we also need to take into account that splitting the flow among a subset $\mathcal{S}(P_v)$ of $|\mathcal{S}(P_v)|$ allowed edges means that the outgoing flow will be divided by $|\mathcal{S}(P_v)|$ on each of these edges.

III. The Notion of Entropic Centrality

Consider the vertex u whose centrality we want to establish with respect to how likely it is that a flow goes from u to any other vertex w. We then look at how many ways the flow goes from u to w through different paths. We define accordingly the probability that a flow starting at u ends at w by

$$p_{uw} = \sum_{P \in \mathcal{P}_{s,w}} \prod_{v \in P \setminus \{s\}} \tau_{P_v}(v) \frac{f(v',v)}{|\mathcal{S}(P_v)|}$$

where the sum is over all paths P from s to w (recall that there is a unique edge going from s to u by construction), (v', v) is the incoming edge of v in a given path P, that is v' is the vertex visited before v in this path, and P_v is the precise upstream of the path P which was traversed to reach v.

Definition 1. The *entropic centrality* $C_H(u)$ of a vertex u is

$$C_H(u) = \sum_{w \in V} -p_{uw} \log_2 p_{uw}$$

where we assume that $0 \log_2 0 = 0$.

A. Non-atomic Entropic Centrality in the Uniform Case

We show next that (perhaps) surprisingly, given a pair (u, v) of vertices, the probability for the flow to go from u to v in the non-atomic split-and-transfer model is the same as that in the atomic transfer model, assuming every subset of edges is chosen uniformly at random, and on every edge, an nth fraction of the incoming flow is outgoing, if the chosen subset of edges for the flow to go out of u is of size n.

Let us first recall a convenient formula.

Lemma 1. *For a positive integer n,*

$$\textstyle\sum_{i=0}^{n-1} \frac{1}{i+1}\binom{n-1}{i} = \frac{1}{n}(2^n - 1).$$

Proof: First we have

$$\tfrac{1}{i+1}\binom{n-1}{i} = \tfrac{(n-1)!}{(i+1)!(n-1-i)!} = \tfrac{n!}{(i+1)!(n-1-i!)n}$$

so that

$$\textstyle\sum_{i=0}^{n-1} \frac{1}{i+1}\binom{n-1}{i} = \frac{1}{n}\sum_{i=0}^{n-1}\binom{n}{i+1} = \frac{1}{n}\sum_{j=1}^{n}\binom{n}{j}$$

and $1 + \sum_{j=1}^{n}\binom{n}{j} = 2^n$. ■

Proposition 1. *In the split-and-transfer case, when a specific branch of the flow carries f_u flow amount to u, each of the $d'_{out,P_u}(u)$ unvisited neighbors receives $\frac{f_u}{d'_{out,P_u}(u)}$ of the flow in expectation.*

TABLE I

Different centrality measures illustrated on Example 1 are found in the left columns. The right columns are obtained from the same example by removing the edge (3,1).

v	$C_{D,in}(v)$		$C_{D,out}(v)$		$C_C(v)$		$C_B(v)$		$C_H(v)$	
1	$\frac{1}{3}$	0	$\frac{2}{3}$	$\frac{2}{3}$	0.44	0	0.16	0	1.6	1.61
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0.44	$\frac{1}{3}$	0.16	0.16	1.6	0.14
3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	0.44	0.44	0.16	0	1.6	1
4	1	1	0	0	1	0.44	0	0	0	0

Proof: Consider a vertex u such that $(u, v) \in E$ and set $n = d'_{out,P_u}(u)$ to be the number of unvisited neighbors of u. There are exactly $\binom{n-1}{i}$ choices containing $i+1$ elements, one of them being v. The total sum of ways to reach v from u is thus $\sum_{i=0}^{n-1}\binom{n-1}{i}$, each of these events is chosen with probability $\frac{1}{2^n-1}$, but also for each term in the sum, the flow involved should be weighted by the number of elements in the subset considered. We thus get that the expected amount $\frac{1}{2^n-1}\sum_{i=0}^{n-1}\binom{n-1}{i}\frac{f_u}{i+1}$ of flow reaches v from u. The above lemma tells us that this simplifies to $\frac{f_u}{n}$. ■

We can thus simplify the probability that a flow starting at u ends at w, obtaining the formula [4], namely

$$p_{uw} = \sum_{P \in \mathcal{P}_{u,w}} \prod_{v \in P} \tau_P(v)$$

where $\tau_P(v)$ is the transfer probability at node v.

The intuition for the notion of entropic centrality given in [4] was that, for a vertex v with high centrality, the uncertainty of where a flow from v would actually terminate would be high, while it will be low for a vertex with low centrality - capturing this uncertainty with the entropy measure. For the split-and-transfer scenario, an analogous interpretation is that a flow originating at a vertex with high centrality spreads more evenly across the graph, while a flow starting at a low centrality vertex is confined and/or concentrated in its spread.

B. An Illustrative Example

Consider the graph shown on Fig.1. If we calculate the probability p_{11} that the flow starts at 1 and terminates at 1, using the transfer model, we obtain $p_{11} = \frac{1}{3}$. In the split-and-transfer model, we have $2^3 - 1$ choices at 1, each with probability $\frac{1}{7}$, out of which 4 cases only contribute to (part of) the flow to 1: (1) $1 \to 1$, indicating the whole flow terminates at 1, contributing to the expected overall portion by $\frac{1}{7}$, (2) $1 \to 1, 2$, and (3) $1 \to 1, 4$, for each half of the flow terminating at 1, contributing $\frac{1}{14}$ in expectation in each case, and likewise (7) $1 \to 1, 2, 4$ contributing $\frac{1}{21}$. Thus the expected volume of the flow that terminates at 1 can be computed as: $\frac{1}{7}\left(1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{3}\right) = \frac{1}{3} = p_{11}$.

Let us assess the right flow instance in Fig. 1 next. The event $1 \to 1, 2$ occurs with probability $\frac{1}{7}$, and only half of the original flow is carried to 2. The split-and-transmit event $2 \to 3, 4$ happens with a probability of $\frac{1}{7}$, and each of the two paths carry $\frac{1}{2}$ of what vertex 2 had received, which is in effect a quarter of the original flow each. Finally, the event $3 \to 4$ happens with a probability $\frac{1}{3}$ and everything vertex 3

TABLE II
AN EXAMPLE OF TRANSACTIONS IN THE BITCOIN NETWORK.

1	Inputs: $\varnothing$ Outputs: 25.0 $\to$ Alice
2	Inputs: 1[0] Outputs: 17.0 $\to$ Bob, 8.0 $\to$ Alice
3	Inputs: 2[0] Outputs: 8.0 $\to$ Carol, 9.0 $\to$ Bob
4	Inputs: 2[1] Outputs: 4.0 $\to$ David, 2.0 $\to$ Bob, 2.0 $\to$ Alice
5	Inputs: 3[1], 4[1] Outputs: 11.0 $\to$ Bob
6	Inputs: 3[0], 5[0] Outputs: 3.0 $\to$ David, 7.0 $\to$ Carol, 9.0 $\to$ Bob
7	Inputs: 6[1] Outputs: 2.0 $\to$ Alice, 5.0 $\to$ Carol

had received from vertex 2 is in forwarded to vertex 4. Thus, from this flow instance vertex 4 obtains $(\frac{1}{7}\frac{1}{2})[\frac{1}{7}\frac{1}{2}+(\frac{1}{7}\frac{1}{2})\frac{1}{3}] \approx 0.0068$ that will contribute to $p_{14}=0.5$.

The entropic centralities of the four nodes are $C_H(1) = C_H(2) = C_H(3) = 1.6121972227$ and $C_H(4) = 0$. [1] If we remove the edge $(3,1)$, the entropic centralities become $C_H(1) = 1.6121972227$, $C_H(2) = 1.45914791703$, $C_H(3) = 1$, $C_H(4) = 0$. Vertex 4 has out-degree $d_{out}(4) = 1$ (counting itself), thus it is an isolated node. Vertex 3 serves as transit node between 2 and 4 once edge $(3,1)$ is removed, it thus has the second lowest centrality. We compare the entropic centrality with other centrality measures in Table I: (1) The *in-degree centrality* $C_{D,in}$ is the ratio of the in-degree (excluding the self-loop) by the number of graph vertices (excluding the vertex itself), the out-degree centrality is defined similarly:

$$C_{D,in}(v) = \frac{d_{in}(v)}{|V|-1},\ C_{D,out}(v) = \frac{d_{out}(v)}{|V|-1}.$$

(2) The *closeness centrality* C_C is defined by

$$C_C(v) = \frac{|V|-1}{\sum_{v\neq u} d(v,u)}$$

where $d(v,u)$ is the shortest distance between v and u. (3) The *betweenness centrality* of a node is the sum of the fraction of all-pairs of shortest paths that pass through it:

$$C_B(v) = \sum_{u,w\in V} \frac{\sigma(u,w|v)}{\sigma(u,w)}$$

where $\sigma(u,w)$ is the number of shortest paths from u to w, and $\sigma(u,w|v)$ is the number of those paths passing through some vertex v other than u,w. If $u=w$, then $\sigma(u,w)=1$, and if $v \in u,w$, $\sigma(u,w|v)=0$.

The raw values of different centrality metrics are not comparable, but the rankings of vertices obtained therefrom are.

IV. APPLICATIONS

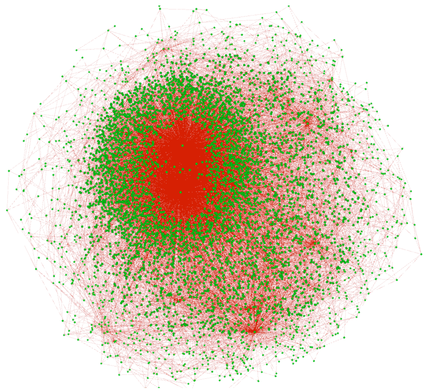
The motivation to consider the split-and-transfer model proposed in this paper is that there are diverse examples of networks for which the flow splits. The network induced by Bitcoin transactions is one such example. In this network, we will assume that vertices are users as identified by their address (this is a simplification since one user may have several addresses). Transactions are happening among users, where a transaction is actually a list of inputs and outputs, as illustrated in Table II. A transaction with no input is a *mining* operation whereby new Bitcoins are created, and the output of the mining operation is assigned to a set of users, possibly including (solely) the miner. In Table II, transaction 1 shows that 25 coins were mined and assigned to Alice. In general, every transaction can possibly include several inputs (which need to be unused outputs of previous transactions) and several outputs. In the network model considered, we deem two distinct users u and v to be connected by a directed edge $(u,v) \in E$ if there exists a transaction where a previous assignment of Bitcoins to u is used as an input to the transaction, and v is assigned some Bitcoins as an output of the transaction. If there are several such inputs between the same users, they are aggregated and shown only once. In our simplified model, we discard the actual amount aggregated over the transactions, and focus on the existence of edges, to reflect connectivity. The graph corresponding to Table II is actually the graph shown on Fig.1 where vertex 1 is Alice, vertex 2 is Bob, vertex 3 is Carol, and vertex 4 is David. For each user, we also assign a self-loop in the graph to indicate the money that is actually not spent - this could either be because the user sent some change back to itself, creating a Bitcoin transaction output, or just because the money was indeed never spent at all (so there is no transaction to itself in the actual list of Bitcoin transactions). The self-loop for David is an example of the later scenario. That David has entropic centrality 0 emphasizes that he is only receiving coins.

Early analysis of the Bitcoin network in terms of in- and out-degree can be found in [7], [8], among with other analysis beyond the scope of this work.

Remark 1. For large networks, computing the entropic centrality can be costly. Thus, when probabilities become very small, it is meaningful to apply a threshold and force the search for paths to stop, and consider that nodes from then on are reached with a probability 0. We used a threshold of 10^{-6} in our experiments reported next.

We took the Bitcoin transaction history for the month of December 2016, and derived the transaction graph among wallet addresses in a manner described above. From the graph obtained, we chose two connected components comprising 5,206 (Fig.2) and 5,251 (Fig.4) vertices to study. We chose these graphs because they are of similar size, a size which is computationally tractable with our current implementation, and also because they have very different structures, as apparent from the figures, where we also report the Kendall-τ coefficient for the ranking of vertices determined by the entropy based centrality with respect to the rankings derived from other (above mentioned) centrality measures. Kendall-τ coefficient is a measure of rank correlation. Higher the measure, the higher the pairwise mismatch of ranks across two lists. A non-negligible value of this distance indicates that

[1] Our Python code for entropy centrality computations on a graph can be found at: https://github.com/feog/Bitcoin-Analysis/



C	$\tau(C, C_H)$
C_B	0.5185
C_C	0.3794
$C_{D,in}$	0.1709
$C_{D,out}$	0.0310

Fig. 2. A component of the Bitcoin network with 5,206 vertices. The table shows the Kendall-τ correlation coefficient between the different centrality measures C and the entropic centrality C_H.

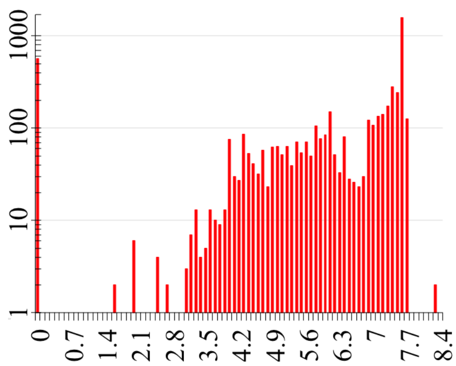


Fig. 3. Entropy centrality frequency (network with 5,206 vertices)

$C_H(v)$	frequency
8.4	1
7	1
6.7	4
2.6	1
2.2	4781
2.1	3
1.5	1
1	1
0	458

C	$\tau(C, C_H)$
C_B	0.3236
C_C	0.4039
$C_{D,in}$	0.2606
$C_{D,out}$	0.0010

Fig. 4. A component of the Bitcoin network with 5,251 vertices. The tables show the entropy centrality stats, and the Kendall-τ correlation coefficient between the different centrality measures C and the entropic centrality C_H.

the rankings obtained from entropy centrality provide distinct information, with respect to the other centrality measures.

Not surprisingly, the out-degree based centrality has the least mismatch, because typically, a large out-degree would help a vertex to spread the flow widely, leading to a high entropy centrality. Yet, it is not always true, since a vertex with small out-degree connected to very few or even one vertex with high centrality would gain high centrality by transitivity. We observe this in our network with 5026 nodes, where the top-5 entropy centralities are approximately 8.26, 8.23, 8.12, 7.68 and 7.62 and belong to vertices with out-degrees 1495, 1494, 12, 265 and 3 respectively. Notice both the fact that the out-degrees are not in a strict order, and hugely different for the vertices even with close entropy centrality. We provide a frequency (log-scaled y-axis) histogram of the entropy centrality (x-axis) in Fig.3. There were 1586 vertices with a entropy centrality approximately 7.6, so we had to use log-scale, and values that occurred only once are not visible and are thus indistinguishable from those which did not occur at all. In this graph, essentially most nodes are well connected to a hub with high spread (there are two such nodes with out-degrees of 1495 and 1494), and thus most of them have pretty high entropy centrality - while 566 leaf vertices have zero entropy centrality. In contrast, for the graph with 5,251 nodes shown in Fig.4 (we embed the entropy centrality (bins of size 0.1) frequency of vertices in the Fig.itself) we see that there are very few vertices with high centrality values (only six of them have centrality of more than 6.6), most vertices have moderate centrality values (they are close to the leaves in the directed graph) and finally there is a large number of leaves which recieve money from several users, without transferring anything forward. In fact, in this graph, six vertices had out-degrees between 429 and 898, and only four more vertices had out-degree more than 1. Every other vertex had out-degree of one or zero. Unsurprisingly, $\tau(C_{D,out}, C_H)$ is very low for this graph, and the out-degree versus entropy centrality based importance of nodes predominantly does not vary.

V. Future Work

There are two obvious future directions of study: refine our model and measure of centrality to take into account the actual amount of the financial transactions in the Bitcoin network (which would require the model to support arbitrary divisions of the flow), and the study of other types of networks (including those that do not conserve volume of the flow, for instance, epidemics) with the current entropic centrality.

References

[1] D. Koschützki, K.A Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, O. and Zlotowski, "Centrality Indices", in Brandes, U. and Erlebach, T. (Eds.), Network Analysis: Methodological Foundations, pp. 16–61, LNCS 3418, Springer-Verlag, 2005.

[2] L. Freeman, "A set of measures of centrality based upon betweenness", *Sociometry*, 40: 35–41, 1977.

[3] S. P. Borgatti, "Centrality and Network Flow", *Social Networks*, Elsevier, 27: 55–71, 2005.

[4] F. Tutzauer, "Entropy as a measure of centrality in networks characterized by path-transfer flow", *Social Networks*, 29, 2007.

[5] A. G. Nikolaev, R. Razib, A. Kucheriya, "On efficient use of entropy centrality for social network analysis and community detection", *Social Networks*, vol. 40, Jan. 2015.

[6] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows. Theory, algorithms, and applications*. Prentice Hall.

[7] F. Reid, M. Harrigan, "An Analysis of Anonymity in the Bitcoin System", *Altshuler Y., Elovici Y., Cremers A., Aharony N., Pentland A. (eds), Security and Privacy in Social Networks. Springer*, New York, 2013. https://arxiv.org/pdf/1107.4524.pdf

[8] Kondor D., Pósfai M., Csabai I., Vattay G. (2014), "Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network". PLoS ONE 9(2): e86197.