Entropic Centrality for non-atomic Flow Networks[1]

Frederique Oggier, Silivanxay Phetsouvanh, and Anwitaman Datta

October 26, 2020

Introduction

- Given a graph to be analyzed, vertex centrality is a notion meant to identify the most important vertices within a graph.
- The closeness centrality gives the highest weight to vertices that 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.

Contributions

- Flow model that allows the flow on a graph to split, based on which they generalize the entropic centrality notion.
- They prove that a case of the entropic centrality gives a new and more powerful interpretation of this centrality measure.
- They provide examples and explore the applicability and limitations of the entropic centrality measure.

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 $E = \{f(u, v)u, v \in V\}$ of edges describes a relation on V.
- Network Flows: Let $d : V \to \mathbb{R}$ be a function such that $\sum_{u \in V} d(v) = 0$. 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
  $d(w) = \sum_{v,(w,v) \in E} f(w, v) - \sum_{v,(w,v) \in E} f(v, w)$ for all $w \in V$.

The Split and Transfer Flow Model

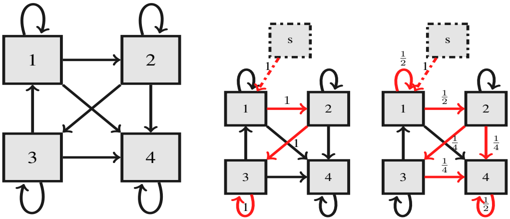


Fig. 1. A small directed connected graph $G = (V, E)$, with $V = \{1, 2, 3, 4\}$, $E = \{(1, 1), (1, 2), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), (4, 4)\}$. The out-degrees are $d_{out}(1) = 3$, $d_{out}(2) = 3$, $d_{out}(3) = 2$, $d_{out}(4) = 1$. Two instances of flow propagation are given on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

Split-and-Transfer and Stop Probabilities

- Now to every vertex v, they attach a probability that describes how the flow that arrives at v will behave next.
- Now, we want the flow to not only to be transferred from one vertex to another but also to be possibly split across different vertices.

Notion of Entropic Centrality

- The probability of a vertex of a flow starting at u and ending at w:
- $p_{uw} = \sum_{p \in \rho_{s,w}} \prod_{v \in P \ \{s\}}^{n} \tau P_v(V) \frac{f(V',V)}{|S(P_v)|}$
- The entropic centrality $C_H(u)$ of a vertex u is:
- $C_H(u) = \sum_{w \in v} = p_{uw} \log_2 p_{uw}$

Illustrative Example

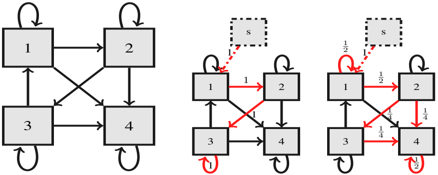


Fig. 1. A small directed connected graph $G = (V, E)$, with $V = \{1, 2, 3, 4\}$, $E = \{(1, 1), (1, 2), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), (4, 4)\}$. The out-degrees are $d_{out}(1) = 3$, $d_{out}(2) = 3$, $d_{out}(3) = 2$, $d_{out}(4) = 1$. Two instances of flow propagation are given on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

- Probability of the flow starting at 1 and terminating at 1 $p_{11} = \frac{1}{3}$.

Illustrative Example

- In the split and transfer model we have $2^3 - 1$ choices at 1.
- Each choice has a probability of $\frac{1}{7}$.
- Out of all the choices, only 4 cases only contribute the flow only to 1.
- The expected volume of flow that terminates at 1 can be computed as:
- $\frac{1}{7}(1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{3}) = \frac{1}{3} = p_{11}$

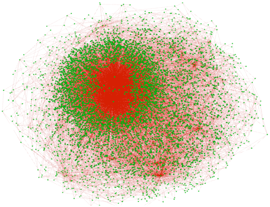
Applications

- There are several examples of networks for which the flow splits. The bitcoin transaction network is one of them.
- In this network, the vertices are users identified by their address, and transactions occur among users. A transaction with no input is a mining operation.

Applications

- The authors took the Bitcoin transaction history for the month of December 2016 and derived the transaction graph.
- They chose two connected components, including 5,206 and 5,251 vertices, to study.

Applications



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.

$C_H(v)$	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.

Applications

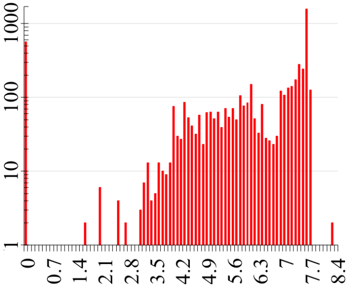


Fig. 3. Entropy centrality frequency (network with 5,206 vertices)

Future Directions

There are two directions for future studies:

- First one is the improvement of the model proposed by authors and measuring centrality to take into account the actual amount of the financial transactions in the Bitcoin network.
- Second is the study of other types of networks.

Conclusion

- The authors generalized the notion of entropic centrality to non-atomic flows.
- This metric could derive exciting results that other centrality measures failed to get.

Questions

- In the first network(figure 2, slide 12), we see a lot of vertices with a high entropy centrality. What do you interpret about the structure of the graph from this data?
- A large out-degree would help the vertex spread the flow widely, which would lead to a higher entropy centrality. Is this argument always true?

The Graph Structure of Bitcoin[2]

Damiano Di Francesco Maesa, Andrea Marino and Laura Ricci

October 26, 2020

Introduction

- In this paper, the authors analyze the bow tie graph structure, attributed to the Web in the seminal paper [3].
- They link the connectivity structure of the Bitcoin users graph to the economical activity of its nodes.

Introduction

- The bow tie structure groups the nodes of the graph depending on their reachable sets of nodes.
- The nodes in the biggest strongly connected component are called SCC.
- The remaining nodes reaching (resp. reached by) the ones in the SCC are called IN.
- The remaining nodes in the biggest weakly connected component are called TUBE, TENDRIL, or FRINGE.
- Other nodes of the graph are called DISCONNECTED.

Introduction

- In the graph proposed by the authors, nodes represent users and edges model the flow of value between such users.
- Nodes of Graph are augmented with their balance, and the edges are weighted according to the Bitcoin value exchanged.
- We have access to the creation dates of the edges, which allow us to perform temporal analysis.

Formal Definitions and Method

- Given the users graph $G(V, E)$ obtained by the clustering heuristic of the Bitcoin transactions graph, the authors assign roles to the nodes in V.

Definition 1. *Given a graph $G(V, E)$, the role of the nodes in V is defined as follows.*

- DISCONNECTED*: nodes not connected to the giant connected component of G^*.*
- SCC*: nodes in the giant strongly connected component of G.*
- IN*: nodes not in* SCC*and able to reach the nodes in* SCC*.*
- OUT*: nodes not in* SCC*and reachable by the nodes in* SCC*.*
- TENDRIL*: nodes not in the previous categories, that either can reach at least a node in* OUT (TENDRILTOOUT) *or can be reached by at least a node in* IN (TENDRILFROMIN)*, but not both.*
- TUBE*: nodes not in* SCC*, that can reach at least a node in* OUT*, and can be reached by at least one node in* IN*.*
- FRINGE*: nodes not in any of the previous categories.*

The nodes having role FRINGE include the ones which are connected to TENDRIL (eventually using an undirected path). For instance, this is the case of a node y reaching $x \in$ TENDRIL, where x can be reached by IN (i.e. $x \in$ TENDRILFROMIN).

Bow-tie Structure

- The authors have used a linear algorithm for assigning the nodes to their corresponding roles.

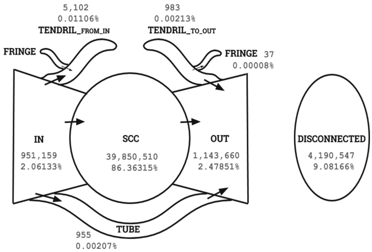


Fig. 1. The bow tie structure of the Bitcoin users graph. For each component is shown its size (i.e. the number of nodes it contains) and the percentage of nodes it contains with respect to the whole graph. For readability the components are not scaled according to their relative sizes.

Bow-tie Structure

- Bow-tie size: The results about the web graph presented in show the IN, OUT, and TENDRIL components all with almost same dimensions(21.29%, 21.29% and 21.52% respectively).
- While the SCC component is slightly bigger (27.74%) and disconnected about one third the size of them (8.24%).
- The SCC component is dominant over all the others.

Bow-tie Structure

- The authors link economical activity of the nodes involved in the different components.
- They intend to show that SCC represents the center of economical activity.
- And they Show that IN nodes move value towards the SCC.
- And OUT nodes correspond to nodes with value credited from the SCC.

Bow-tie Structure

- To better understand the type of nodes with different roles, they evaluated the following metrics with respect to each role.

  - AddrNum: counts the number of addresses in a cluster. This measure can be used to indirectly estimate the cluster activity.
  - Balance: measures the current (at data collection time) balance of a cluster.
  - ValueRec: expresses the total value received during a cluster lifespan, or up to the time of data collection if the cluster is still active. It is defined as the sum of all payments received.
  - TransIn: represents the number of payments received by a cluster (coinbase rewards included). This measure is useful in estimating the economical importance of a cluster in the way of collecting payments.
  - TransOut: measures the number of payments done by a cluster. It is computed as the number of transactions originated from a cluster. This measure represents the economical importance of a cluster as its weight in issuing payments. Note that both TransIn and TransOut measures the in- and out-degrees of the clusters, including self-loops and multi-edges. These kind

Bow-tie Structure

- Coinbase Transactions

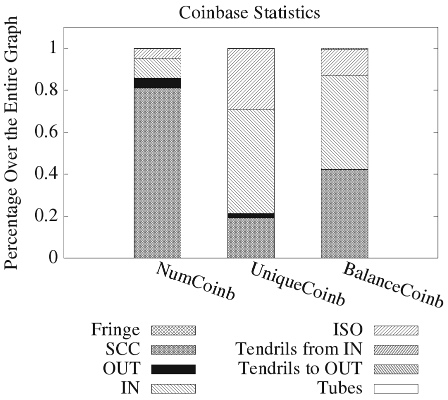


Fig. 3. Fraction of NumCoinb, UniqueCoinb and BalanceCoinb for each role component over the entire graph.

Temporal Analysis

- A clear advantage of building a graph from historical data is that it would make it possible to perform a temporal analysis of the graph structure and to study the evolution of its components.

Temporal Analysis

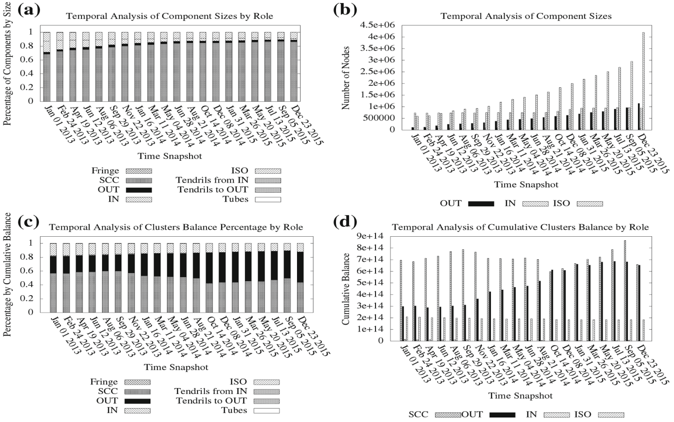


Fig. 4. *Temporal analysis* of role component size percentages over the whole graph (a), temporal analysis of component sizes restricted to the three most significant components excluding the dominating SCC only (i.e. OUT, IN and DISCONNECTED) (b), temporal analysis of cumulative current balance percentages over the whole graph (c) and temporal analysis of cumulative current balance of the three most significant components (i.e. SCC, OUT and DISCONNECTED) (d). For the precise GMT time snapshot dates, see [6].

Future Directions

There are two directions for future studies:

- The economical meaning of the nodes in the different components of the bow-tie can be further investigated by exploiting more sophisticated deanonymization techniques.
- The authors are currently performing the same analysis for the graph obtained from the Ethereum blockchain, with the goal of comparing the economies of the two cryptocurrencies.

Conclusion

- The authors presented an analysis of the structure of the Bitcoin users graph obtained from the Bitcoin blockchain.
- They investigated the economical meaning of the different components of the graph.
- They also presented a temporal analysis by considering the evolution of the graph over time.

Questions

- We saw that the nodes in the SCC have the highest average value for the measures introduces by the authors(Figure 3 slide 26). what does this data represent in terms of economic activity?
- What does the evident difference between the first two columns in figure3(slide 26) tell us?

References

F. Oggier, S. Phetsouvanh, and A. Datta.
Entropic centrality for non-atomic flow networks.
In *2018 International Symposium on Information Theory and Its Applications (ISITA)*, pages 50–54, 2018.

Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci.
The graph structure of bitcoin.
In *International Conference on Complex Networks and their Applications*, pages 547–558. Springer, 2018.

Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener.
Graph structure in the web.
Computer networks, 33(1-6):309–320, 2000.