Implications of Network Structure of the SAP Online Knowledge Community Platform

- **Vigneshwaran Giri Velumani**

The SAP Network has 6090 Edges & 3415 Vertices. After simplifying the Edges count drops to **4120**. So basically, we are simplifying the network by combining the counts of answers provided by each User ID. Refer Fig.1

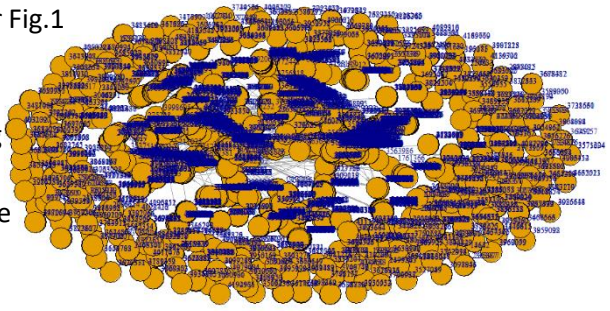


Fig.1 Plot of Entire SAP Network Community

Reciprocity:

The very less reciprocity value of **0.005825243** shows that the whole network is heavily unidirectional which is understandable because in any knowledge sharing community forum/platform there will be very less people who have more knowledge than others in multiple domains. It also signifies that most of them are not mutually linked with each other except for the very few active responders of question.

Transitivity:

A low transitivity value of **0.009985725** which is **399429** out of **40000000** indirect connections are transitive. This could be interpreted in our scenario like clustering happens only within groups of people in the same domain and only a very few active responders serve as the common nodes in the entire network.

Average Path Length: 3.983 | Diameter(Regular Weights Method):26 | Diameter(Inverse Weights Method):14.27

The diameter of **14.27** seems reasonable for such a large SAP knowledge sharing community network considering the number of nodes and edges in the network.

Cliques and their Structure:

Cliques of Size 5:

There are **5** Cliques of Size **5**. If we can refer to Fig.2 nodes like **983891**, **2704623** have more inward arrows which signifies that these are the User ID in the clique which received the responses from multiple users for the question threads they created.

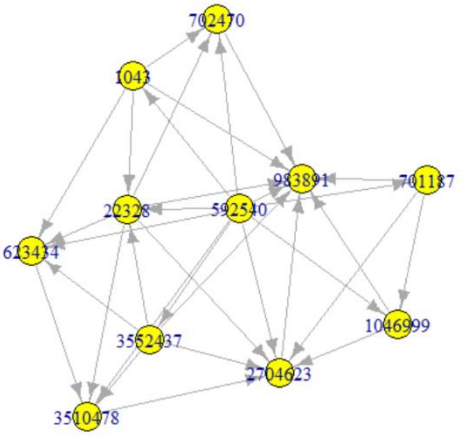


Fig.2 Plot of Cliques of Size 5

Cliques of Size 4:

There are **39** Cliques of Size **4** as shown in Fig.3.From the figure we can see that multiple inward arrows present for **338884**. May be this is one of the threads where multiple User IDs are trying to seek knowledge from. Among the nodes reaching out to **338884**, we can see that node **3583224** is also trying to access information and pass it on to its neighboring nodes. This is one of the examples of strength of weak tie. And the node **338884** can be considered as one of the structural hole access nodes in this network.

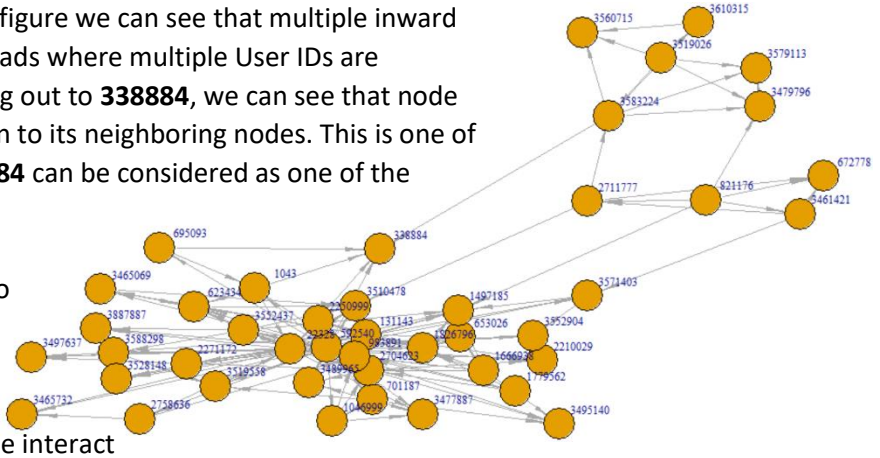


Fig.3 Plot of Cliques of Size 4

Cliques of Size 3:

There are **335** cliques of Size **3** as shown in Fig.4. This also shows the strength of the triadic closure which is **335**. So, among each of these **335** cliques the triadic closure is strong, but the triadic closure of the whole in general is not that strong. In a knowledge sharing community people interact within themselves in small clusters where knowledge sharing is high, but only within that cluster. So, there are only a few know ledged experts who link with these triadic closure communities and serve as the source of new, non-redundant information.

Cliques of Size 2:

The network has **3320** Cliques of Size **2**. For a network having a size of **3415** Nodes, having **3320** cliques of Size 2 is common. It also signifies the fact that each user is at least connect with either one other user or one other thread. So, the remaining 95 nodes can be either dormant users in the network who did not interact with any other user or did not post a question thread to ask questions/ seek help from other users in the community.

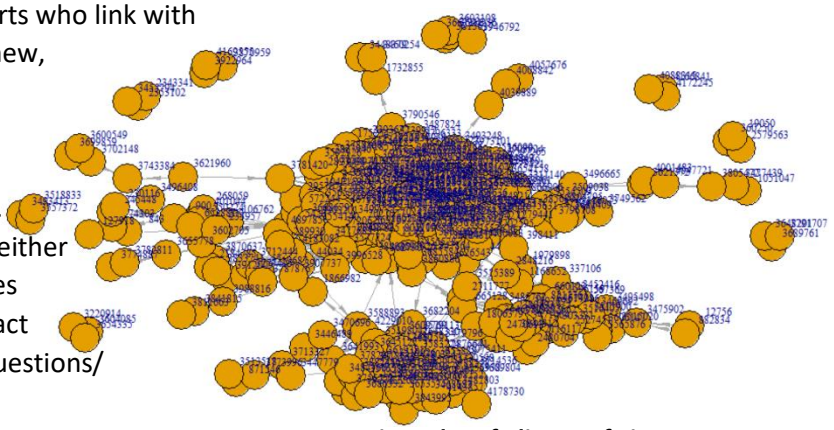


Fig.4 Plot of Cliques of Size 3

Embeddedness/ inverse of structural hole access: Fig.5

To find the nodes which act as the structural holes in this network we must look at the nodes which have the lowest constraint Value. So, by this definition the top 5 nodes which act as structural holes are: **592540, 821176, 3583224, 131143, 44034.** According to our network nodes with high Structural holes or Lowest Embeddedness shows that the ideologies and knowledge of these people are unique and have very little overlap with others. So, in this knowledge sharing community they act as nodes that transfer novel information among many clusters in the communities in the whole network. The 10 blue square nodes in the fig5 are the nodes having highest Structural hole access

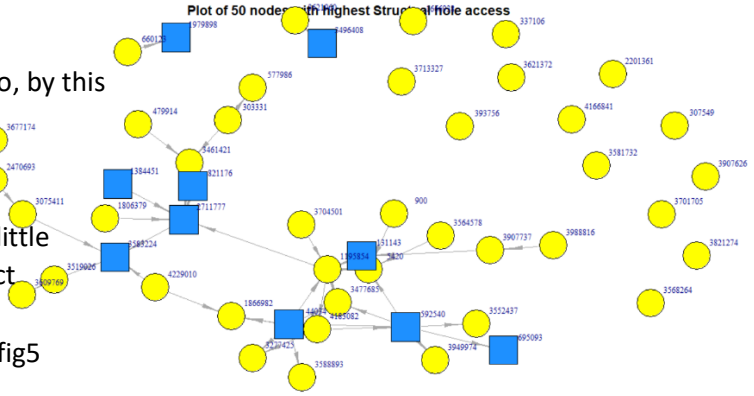


Fig.5 Plot of Nodes Having highest Structural Hole Access

Degree Centrality: Fig.6

Since Degree centrality measures the counts how many neighbors a node has, especially in a directed network, we consider in-degree which is the number of in-coming links. So, the more the degree centrality value of a node, the more is its interaction with its neighboring nodes. If a person answers a lot of questions related to different topics or creates several threads to help people gain more knowledge by hosting several Q&A sessions, then that node's(person's) Degree Centrality value will be very high. In our case, the following 5 nodes have highest degree centrality: **592540, 821176, 131143, 3583224, 22328**. The **10** blue square nodes in the fig.6 are the nodes having highest Structural hole access. For the sake of simplicity of the visuals, the neighboring nodes of these highest structural hole access nodes have been ignored

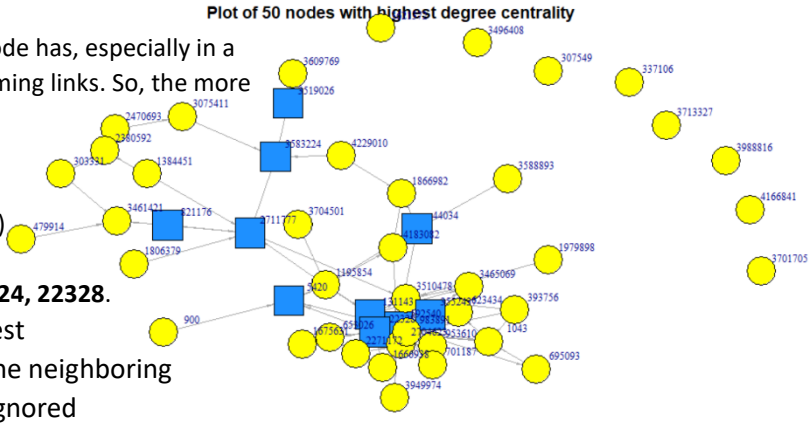


Fig.6 Plot of Nodes Having highest Degree Centrality

Node Betweenness: Fig.7

Betweenness centrality measures the extent to which a vertex lies on paths between other vertices. Vertices with high betweenness may have considerable influence within a network by virtue of their control over information passing between others. They are also the ones whose removal from the network will most disrupt communications between other vertices because they lie on the largest number of paths taken by messages. So according to the above definition, we can find that 5 of our nodes :**1195854,4183082,592540,44034,2711777** have the highest betweenness centrality which means they control majority of the information exchange in the network, also, removal of them from the network might disrupt the communications in the whole network. The node 1195854 which is the node having highest betweenness has been indicated using blue color in the Fig.7

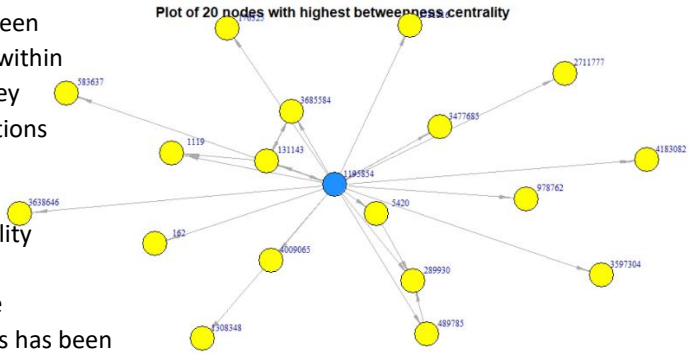


Fig.6 Plot of Nodes Having highest Degree Centrality

Edge Betweenness:

Similarly edge betweenness is a centrality measure using the edges where majority of the information flows through these edges. If we remove these edges from the network then it might disrupt the communication in the whole network. Following are the 5 edges having the highest edge betweenness centrality in our SAP network: **2153,10,1272,8,769**

Local clustering coefficients – Transitivity:

Transitivity refers to the extent to which the relation that relates two nodes in a network that are connected by an edge is transitive, which implies that each component is a clique. So, it measures how strong the cluster is created in those cliques. In our SAP network, any node which has transitivity = 1 means that those nodes are deeply embedded in a cluster that they share very similar ideas and it is very rare that someone in this cluster will think different.

Edge Weight Vs Average Betweenness: Fig.7

From the left Plot of Fig.7 we can observe that lower Average Betweenness of edges values are observed for lower values of Edge weights, but at the same time, we can also observe a few Average

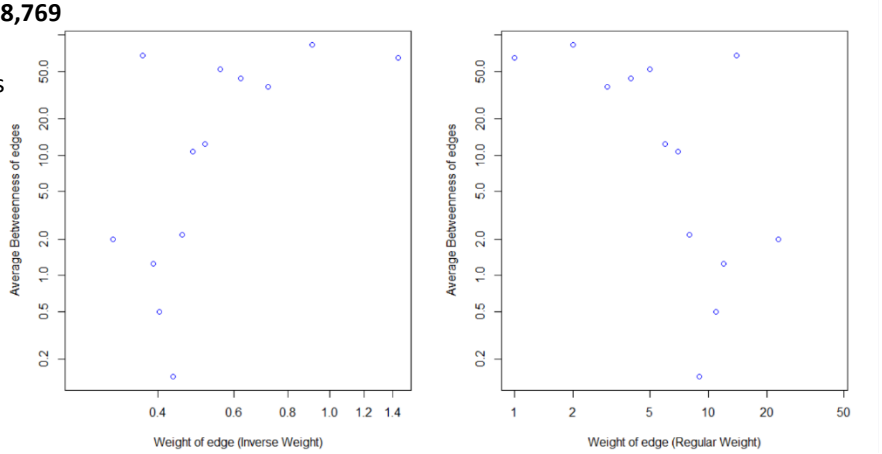


Fig.7.Edge Weight Vs Average Betweenness

Betweenness values exists for lower values of Edge weights as well.

But the overall plot can be generalized as direct relationship with the increasing edge weights, the average betweenness value also increases linearly with a positive slope. From the right plot of Fig.7 we can observe that higher Average Betweenness of edges values are observed for lower values of Edge weights, but at the same time we can also observe a few Average Betweenness values exists for higher values of Edge weights as well. But the overall plot can be generalized as inverse relationship with the increasing edge weights, the average betweenness value decreases linearly with a negative slope.

Clustering Vs Average Betweenness: Fig.8

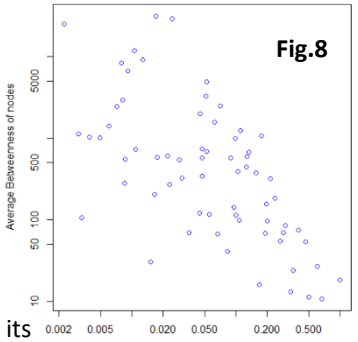


Fig.8

Using the above definition and combining them together, it can be understood that nodes which have fewer clustering values have the highest betweenness because those people who are not tightly connected to any cluster serve as the structural hole and act as the main node for information flow through them. Moreover, the plot also proves this inverse relation between the average betweenness and the clustering coefficient values that nodes which have fewer clustering coefficients have higher average betweenness and the relationship is linear with a negative slope.

Degree Vs Average Betweenness and Clustering & Constraint: Fig.9

Betweenness Centrality and Degree centrality measure almost the same thing - the importance of the node to its neighbors, except degree measures the number of incoming links to itself whereas betweenness measures how important is the node for information exchange in the network. A node may have very high betweenness, but that does not necessarily mean that it should have a high degree and vice versa, but in genera; if we aggregate the values for mean, we can find from the plot that there is a direct relationship with the linear line with positive slope.

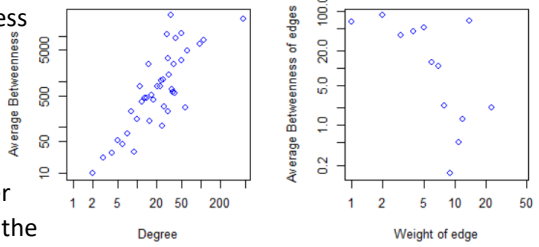


Fig.9

Clustering coefficient is a measure to denote how tightly knotted the nodes are in their cluster whereas degree denotes how many neighbors both internal and external nodes of the cluster the node is in are connected to it. So, any cluster of people sharing similar ideas in the same domain will have high clustering coefficient, whereas less degree. This inference is evident from the plot that with increasing degree centrality, the average clustering coefficient value decreases showing an inverse relationship with a linear line with a negative slope. The more knowledgeable people are usually located outside those clusters and connected to nodes from different clusters.

Embeddedness denotes absense of new ideologies and high overlap of thoughts and ideas of nodes in a cluster. So, a node having high degree will have less embeddedness. This is evident in the plot that with increasing degree centrality, the Embeddedness coefficient of the nodes decreases showing a negative relationship with the regression line having a negative slope.

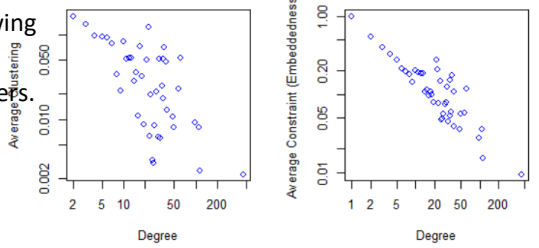


Fig.10

Executing community detection algorithm: Walk trap

From the Fig.11 we can see from the colors that we have multiple different communities in this network.

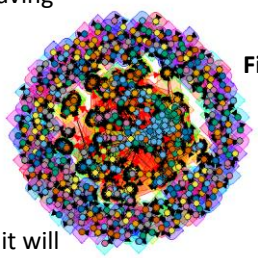
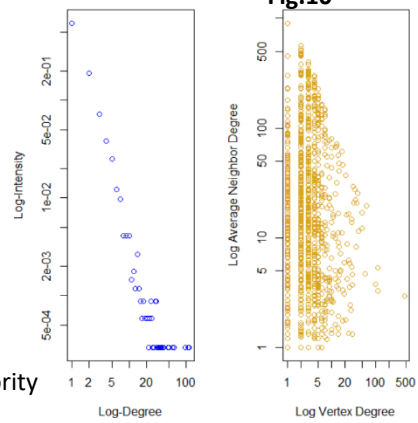


Fig.11

Strength of Nodes: Fig.12

Strength is a weighted measure of degree that considers the number of edges that go from one node to another. In this network, it will be the total number of interactions of each character with anybody else. In SAP network this means that these top 5 contributors have been answering to the questions(edges) threads created by others.



Authority Fig.13

Authority score is another measure of centrality initially applied to the Web. A node has high authority when it is linked by many other nodes that are linking many other nodes. So in SAP network high authority nodes in blue denote that any new and innovative ideas originate from someone, these high authority nodes make sure that the reach to other people in the network. Fig 13

Ego Network for node having highest Degree:

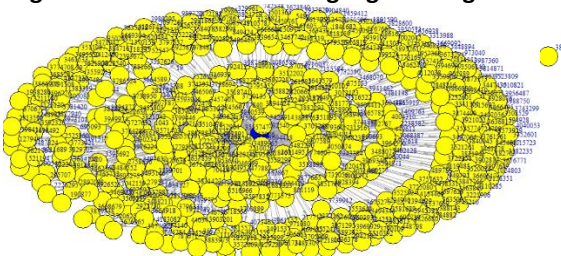


Fig.14 Ego Network of Node-592540

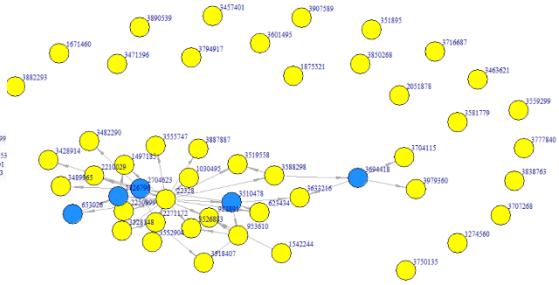


Fig.13 Plot of 50 Nodes having highest Authority score

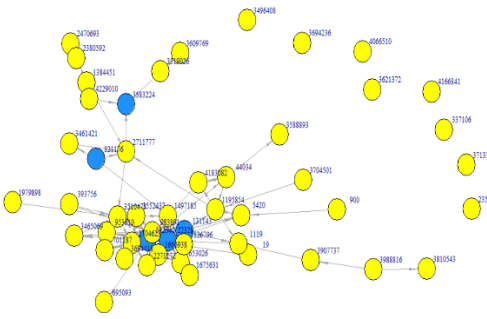


Fig.12 Plot of 50 Strongest Nodes

Conclusion: After analyzing all the metrics it is evident that majority of members are not directly connected with each other but broken down into several clusters, but it's the weak question thread that ties them together. Many members have knowledge only about their domain, and the unique knowledge sharing occurs through (weak ties) the people outside these clusters who have knowledge in multiple domains.