1. Research Question(s):

Social network visualization is becoming an important aspect for organizations in order to discover new insights. Such explored phenomena could help the organization in better resource management. My research question is encouraged by how a start-up can utilize such employee's communication data for deriving important decisions.

- Are there any observable patterns in the way individuals communicate and collaborate in an organization?
- Who should be approached for circulating information in an efficient manner?
- Can we identifying important individuals in an organization?
- In which channels/projects is the individual most active?
- What is a typical activity pattern of an employee over a day?
- How employees interact in an overall slack public channel versus in any one particular public slack channel?

These research questions have derived the design choices for visualization. Later on the visualization is evaluated on the basis of its ability to answer these research questions.

2. Data Source(s):

Slack Public Channel Data containing 11714 messages between 99 individuals over 10 slack channels is created to explain the concept (synthetic data).

The data is preprocessed in Python to avoid complicated parsing in Processing. 11 graph files (.GEXF) were generated using Gephi with all basic network metrics. Although a CSV file could be extracted from Slack API but is not done here to avoid data confidentiality and privacy.

3. User Instructions:

On running the sketch, initial screen shows the overall slack communication network of how employees interact based on all the public channels. Further details would be displayed once the user clicks on any node as shown in Figure 1.
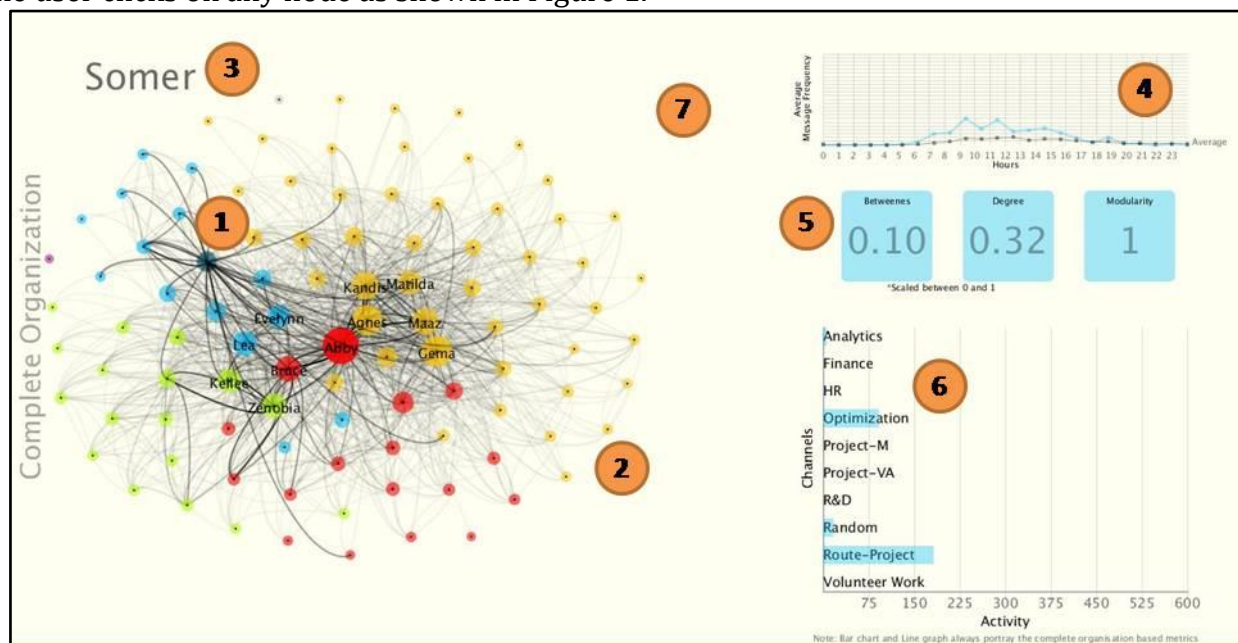

Figure 1: Sketch after the selection of node

1. Once the node is selected it highlights with grey color and show its links in black. Elements 3-6 show the details corresponding to the selected node.
2. It shows a network based on all slack channel communication, user can select on any other node in the network and elements 3-6 would automatically update.
3. The label of the node is shown (in our case it corresponds to employee's username).
4. Line graph shows the variation of average message frequency over a 24 hour period. Blue line shows the selected node's message frequency and grey line is the average frequency line. On hovering a mouse over the blue line, x and y axis values of that point would be displayed on top of the line chart.
5. Basic network metrics (betweeness centrality, degree centrality and modularity) are shown. This is also embedded in the visuals and would be explained in visual encoding section.
6. Bar chart shows the activity strength of the selected node in different slack channels. If a mouse is hovered over any blue bar a list would be shown of all the people the selected node has communicated with, in that particular slack channel, as shown in Figure 2A. Selecting any blue bar would open the node-link diagram of that particular channel as shown in Figure 2B.
7. Clicking anywhere on the background would clear the selection.

Figure 2A: Left-Selecting Optimization Bar in bar chart, Figure 2B: Right- Network of selected bar
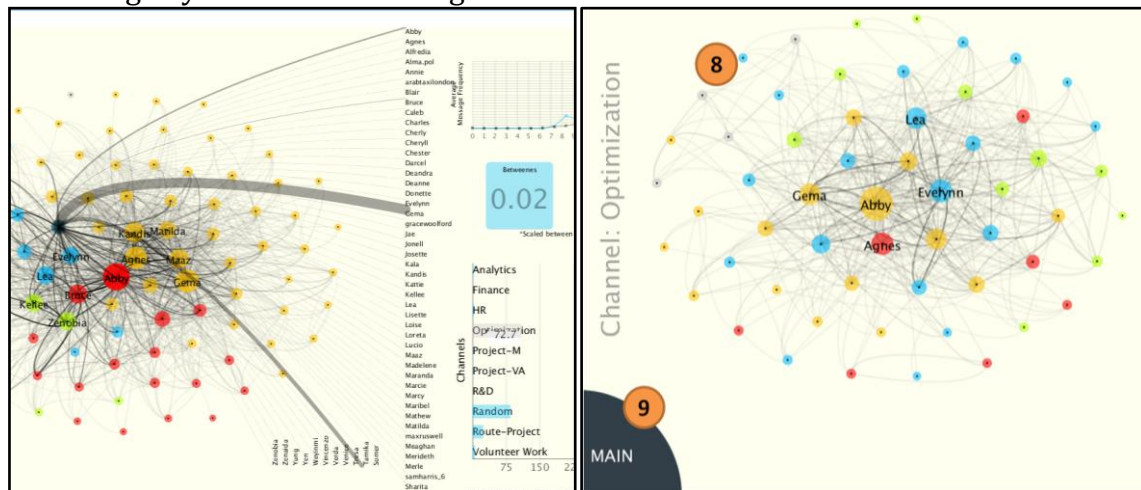
8. A network for selected channel is shown only; selecting any node would display elements 3-6. However, user must be aware element 4 and 6 display the statistics based on all channels, not for just the selected one.
9. Clicking on this element would take back to the initial sketch that has the complete network.

4. Design Justification:

4.1 Visual Encoding:

Colors: Ivory color was used for the background because it is not as bright as white, yet gives a canvas like feel. A user can see the different colors and opaqueness distinctly over a light background. The nodes belonged to different clusters making them categorical while on the other hand had two important continuous attributes. Stevens color chart shown in lecture 3 was utilized to inform color selection choices. (Gramazio et al., 2017)'s Colorgorical was also tried for generating colors with different settings. It influenced some decision but aesthetic appeal and slack's logo color was also considered before finalizing the node's color. Various

colors were tried for the links between nodes, only black and its shade were suitable, clear and easy to follow against ivory background. The strong and weak links were easy to distinguish because of the high brightness variation offered my black color. The whole application follows a similar color theme as slack's logo color and is nodes are bright enough to capture user's attention (Kirk, 2012).

Symbolization: Two important attributes about the network are betweeness centrality and degree centrality; and both were required to be embedded in visualization to give clues about the network structure without overloading the working memory of the user. (Zhu et al., 2010)'s approach was used to incorporate these insights in nodes using color opaqueness and node size for betweeness centrality and degree centrality, respectively. Visually encoding this data instead of presenting in tabular form gives the benefit of using only perceptual cognitive process as stated by Zhu.

Curves were used for drawing edges instead of a line by utilizing the code from Network-lecture 9. A study by (Huang et al., 2016) suggested that using curves for drawing link increases readability and visual aesthetic pleasantness of the graph. Moreover, it was also learnt that the shape of the curve showed the direction without cluttering (Wood, 2012). The straight end of the curve is the origin and the curvy end of the curve is where the edge terminates. In order to avoid hairball affect in the network that has thousand of edges, the stroke weight was kept small and the link strength was shown using the color intensity. Since the research question asks us to look for important actors, weak links were made out of focus my reducing their color's opaqueness. It helped tremendously in avoiding the hairball problem. The techniques were adopted from Zhu's research and Network Lecture 9.

Line graphs are an effective method for showing variation of quantity over time (Kosslyn, 2006). Thus, it was selected to project the message frequency of the employee over a day. To emphasize on selected node's line in the chart, a larger stroke weight and a bright color is chosen while a dull color is chosen for the average message frequency line. Moreover, no data labels are shown to avoid cluttering in the chart unless user interacts with the chart. In order to make the two lines (actual node line and average line) distinguishable, points on the lines were marked by discriminable symbols (circle and square, respectively) as shown in Figure 4. All techniques were adopted from (Kosslyn, 2006)'s line graph principles.

Layout:
(Zhu et al., 2010)'s study demonstrated that using multiple integrated windows or screen is ineffective and makes it hard for the analyst to consolidate all the information in the memory. Thus, this visualization is based on a single window displaying all relevant information; however, this is not always possible since the display capacity is limited.

Among various techniques to visualize network, (Zhu et al., 2010)'s research and (Munzner, 2014) Chapter 9 claim that node-link does the best job in visualizing a network, if structure is of importance. The research questions we raised relied heavily on the network structure in order to find the pattern of collaboration. This forced us to use node-link diagram, even though it is known for causing 'hairball' problem as we observed in session 9. Idiom used for laying out the nodes was Fruchterman Reingold-force directed using Gephi because it creates a compact yet sparse network (Combe et al., 2010). Some approaches aligned with Zhu's research were also employed as mentioned in symbolization section for avoiding the 'hairball problem'.

4.2 Interaction:

A study by (Sundar et al., 2014) found that mouse over encouraged greater engagement from users in compare to other forms like clicking, zooming and scrolling. However, algorithms slow response in updating charts while hovering mouse over a network forced us to take a click based approach. Munzner states that "users are unwilling or unable to wait" if the interaction takes more than a fraction of seconds to respond. Additional detail on demand was fostered by mouse over, where possible, to encourage user engagement.

The interaction and layout is built loosely around (Shneiderman, 1996) visual information seeking mantra. The overview of the whole organization's communication network is shown using a node-link diagram. User is suggested to interact by clicking on the node which shows further detail including a line chart and a bar graph. The bar graph shows over view of the activity over different slack channels of the selected individual. If mouse is hovered over it, the message frequency is shown over that channel and node's connectivity with other employees in that channel is shown. This detail could suggest more drill down and on clicking that bar a network of selected slack channel is shown. The cycle of interaction continues till user has formulated the answers.

Figure 3: Network showing important individuals

5. Data Insights:

The first screen explicitly showed the key players in the organization due to large node size, high color opaqueness and node's position near the centre of the graph. Abby being the president of the organization is quite noticeable. Other key players in the organization could also be spotted easily eg: Evelynn, Kandis etc as shown in the Figure 3.

A pattern can also be observed in the network because of different subgroups being highlighted with different color on the basis of modularity metric calculated in Gephi. Inspecting further shows that nodes belonging to a particular cluster communicate with each other over a particular slack channel.

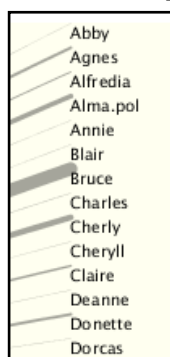Figure 5: Abby's message activity over a typical day

Figure 4

Strong links can be observed between Bruce and the other red node (Donette). Such strong collaboration could help in forming teams in the future that is compatible. One can also observe from Figure 1 that Somer is only active in three channels and such individuals should be encouraged to get involved in other projects to broaden their portfolio and skill-set. Figure 1's line chart also tell us that Somer is active only during office hours and has a good work life balance. While if we see Figure 5, we can observe that Abby being the founder of start-up has little work life balance and is even messaging after office hours. One can also observe the pattern of how Abby communicates in different slack channel, like in analytics Abby is collaborating the most with Bruce as shown in Figure 5.
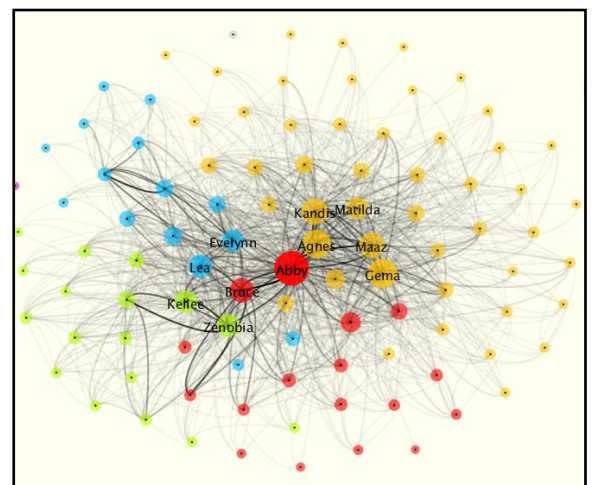
6. Critical Evaluation and Reflection (Postgraduate students only):

Data insight has answered the research questions raised but the success of finding a pattern in a network is hard to evaluate from the design since it is a broad term. All specific and explicit research questions can be easily answered as shown in Section 5. A pattern can also be seen by the node color which shows different sub groups within a network.

However, network changes when moving from the whole organizational view to any one particular slack channel could be improved using animation. The dynamic network that shows how network has evolved over time would also be more helpful in pattern detection, currently, the design has limited or no coverage of time domain. Cluster color for the nodes is not consistent in all slack-channel networks that can cause confusions for a user.

The design is not scalable and increasing the number of nodes could cause clutter in the node-link diagram. A better approach would be to use multi level force directed idiom as suggested by Munzner which can handle thousands of nodes and links. A study by (Henry, 2007) claims that combining node-link diagram and matrix view enhances the network exploration experience of a user. This could be considered in future designs.

Figure 6: Mixed no cross offers best readability

Finally, the curved link is better than lines however, (Huang et al., 2016) work also suggests that curves with no crossing and a mix of lines (mixed-nocross) offers the best readability, shown in the Figure 6. Although mixed-nocross would be hard to implement and is expected to be computationally expensive but offers enhanced network readability and can be tried in the future.
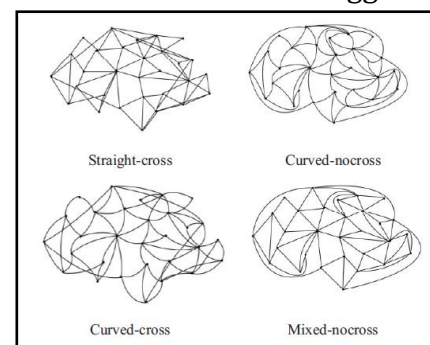
7. References

Combe, D., Largeron, C., Egyed-Zsigmond, E. & Gery, M., 2010. A comparative study of social network. *International Workshop on Web Intelligence and Virtual Enterprises* , 2.

Gramazio, C.C., Laidlaw, D.H. & Schloss, K.B., 2017. Colorgorical: creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics*.

Henry, N.a.F.J.-D.a.M.M.J., 2007. NodeTrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics*, 13, pp.1302--1309.

Huang, W., Peter, E., Seok-Hee, H. & Henry, B.-L.D., 2016. Effects of Curves on Graph Perception. *Pacific Visualization Symposium (PacificVis) IEEE*, pp.199-203.

Kirk, A., 2012. Conceiving and Reasoning Visualization Design Options. In *Data Visualization: a successful design process*. Packt Publishing Ltd. Ch. 4.

Kosslyn, S.M., 2006. Graph Design for the Eye and Mind, Creating Line-Graph Variants and Scatterplots. Oxford Scholarship Online. pp.142-45.

Munzner, T., 2014. Arrange Network and Trees. In *Visualization Analysis and Design*. CRC Press. Ch. 9. pp.201-18.

Shneiderman, B., 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings IEEE Symposium on Visual Languages*.

Sundar, S.S. et al., 2014. User Experience of On-Screen Interaction Techniques: An Experimental Investigation of Clicking, Sliding, Zooming, Hovering, Dragging, and Flipping. *Human–Computer Interaction*, 29(2), pp.109-52.

Wood, J., 2012. *London Moves, visualising 16 million cycle journeys across the UK capital, TEDx , YouTube*. [Video] Available at: https://www.youtube.com/watch?v=FaRBUnO5PZI [Accessed 15 April 2017].

Zhu, B., Watts, S. & Chen, H., 2010. Visualizing social network concepts. *Decision Support Systems* , 49(2), pp.151-61.