SpaceBook

A Case Study of Social Network Analysis in Adjacency Graphs

Gabriel Wurzer[1], Wolfgang E. Lorenz[2]
[1,2]TU Wien
[1,2]{gabriel.wurzer|wolfgang.lorenz}@tuwien.ac.at

In this paper, we have adopted methods from Social Network Analysis in order to analyze adjacency graphs. Our intent was to uncover as much hidden structures as possible so as to improve adjacency requirements before they are used further on during the design process. To that end, we have conducted a case study using two readily available software packages (Gephi, Pajek), concluding that these could benefit from being more transparent about the underlying algorithms and more geared towards the problem domain 'adjacency analysis' when it comes to data entry and visualization. As a matter of fact, we produced an open-source prototype called SpaceBook, which customizes computation and visualization in the aforementioned spirit.

Keywords: *Adjacency Graph, Social Network Analysis*

INTRODUCTION

In architecture, it is usual to prescribe functional relationships in linguistic terms, e.g. '*dining should be close to cooking, but cooking should be far from sleeping*'. More formally, each of these statements can be seen as a pair of functions or spaces having a prescribed *relation* (e.g. mandatory, desirable, neutral or negative; interaction or no interaction) which can be mapped to numbers (mandatory=1, desirable=0.5, neutral=0, negative=-1; interaction=1, no interaction=0). The set of all relationships establishes an (undirected, weighted) *adjacency graph* in which each *node* stands for a space or function and each *edge* for the adjacency relation between the nodes it connects.

Numerous authors have shown that such graphs can be used to drive space layout. Surprisingly, one thing that that has not yet been covered is how to analyze that graph itself! "*Would it be possible to gain insight into the prescribed adjacency relations themselves, using off-the-shelf methods from Social Network analysis?*" was thus our motivation for conducting this case study. In more detail,

1. we have looked into several metrics for analyzing an adjacency graph theoretically, all of which are reported in the 'Background and Related Work'. After doing that, we
2. went on to try to apply this knowledge using two of the most prominent software packages used for SNA (see 'Case Study'), only to fail in several aspects having to do with practicability for daily work (see 'Discussion'). Instead of dismissing the subject as being 'beyond architectural practice', we
3. implemented a quick mash-up between Excel and NetLogo (Wilensky 1999), which is easy to

use and comprehend even when not coming from SNA (see 'SpaceBook'), available for free at [3].

Beyond conducting the case study, we would summarize our contribution as follows: Our occupation shows that there needs to be a mapping between adjacency relationships (used by architects) and weights (used by SNA/graph analysis). This mapping is not always the same, it needs to be adapted as stated in the section 'SpaceBook', in order to feed the used algorithms. Secondly, we find that there is disagreement between the notions of importance (more correctly: of centralities) that the SNA community itself uses, and thus design computing needs to have access to the underlying algorithms in order to justify what is being computed [which, in our case, is done via the use of an open-source paradigm].

BACKGROUND AND RELATED WORK

Centrality refers to the importance of each node in a graph when compared to every other node. *Centrality measures* have been investigated first as part of graph theory and later in network analysis. In architecture and urban planning, the most prominent example of their use lies in the *Space Syntax* model (Hillier and Hanson 1984) which thereby determines the prominence of each element within a set of intersecting line segments (*axial lines*) which are transformed into a graph (*justified graph*; see Wang and Liao 2007 for a detailed description and implementation). Space syntax commonly uses *betweenness* and *closeness* centrality (for others, see e.g. by Porta et al. 2006; Law et al. 2012):

Betweenness centrality. The shortest path from each node to every other node is computed (Freeman 1977). For each node lying within the shortest path (i.e. all nodes between start and end node), we add 1 to the node weight. Nodes with a high total weight are thus highly prominent from the standpoint of transitioning between nodes.

Closeness centrality. For every node, we compute all shortest paths to every other node. For each shortest path, we add the number of nodes in that path to the starting node's weight. This measure gives the "farness" of every node, of which the reciprocal value is the "closeness" we are after. Nodes with high closeness are central hubs from which it is easy to reach other nodes.

In case that edges signify "roads", we might add the edge length instead of 1 in the above computation. Other centrality measures include:

Degree centrality. Each node counts its number of edges, giving a measure of "connectedness".

Eigenvector centrality. An extension to degree centrality which assigns a high weight to a node if it is connected to nodes having a high weight themselves. In more detail (Bonacich 1972; Bonacich 2007), (1.) a score of 1 is attributed to every node, (2.) the scores of each node is recomputed as a weighted sum of centralities of all nodes its neighborhood, (3.) the score is normalized by dividing it by the overall maximum score, then (4.) steps 3 and 4 are repeated until the score stop changing.

This is a selection of the most common centrality measures which is far from complete. Further examples include *Random Walk Centrality* (Noh and Rieger 2004; see Fidler and Hanna 2015 for an application in space syntax) and *Information Centrality* (Stephenson and Zelen 1989). The study of networks not solely from the viewpoint of the graph being analyzed but with a view towards the position of each individual actor's position within that network has furthermore led to the umbrella term *Social Network Analysis* (SNA; Borgatti et al. 2009), which furthermore includes the following measures:

Density. A ratio defined by a node's degree versus the number of possible edges (these are $n(n-1)/2$ in a complete, undirected graph).

Clique analysis. Identifies fully connected subgraphs (cliques), i.e. groups of nodes that are highly dependent of each other.

SNA also seeks to characterize a graph by its tendency to cluster (Holland and Leinhardt 1971; Watts and Strogatz 1998), to be measured by:

Local clustering coefficient. A ratio defined by the number of edges between a node's neighbors

versus the total number of possible edges between its neighbors. This measure gives an indication of embeddedness of a node.

Average local clustering coefficient. Measures the overall clustering of a graph by calculating the mean of each node's local clustering coefficient.

Global clustering coefficient. Overage clustering of a graph based on the determination of triplets: For each node, one examines each pair of neighbors. If that is connected, we have a closed triplet, if not, an open triplet. Overall clustering is then simple the number of closed triplets divided by the total number of triplets.

A high clustering coefficient is typical for small-world graphs, where every node can be reached from every other node by a small amount of hops (e.g. 6 degrees of separation in real-world social networks).

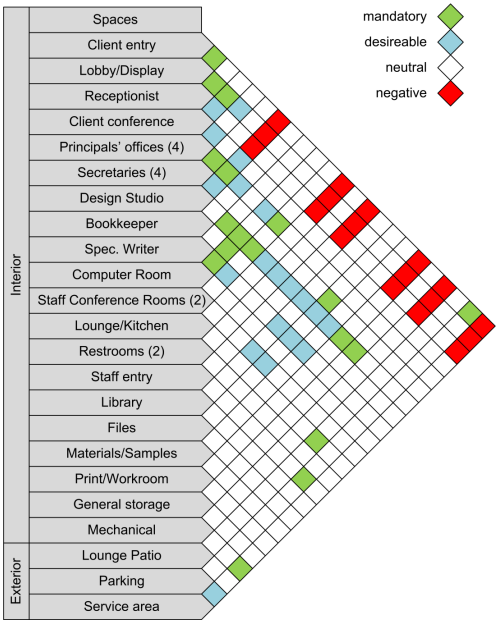


Figure 1
Adjacency matrix of a design studio from White 1986, pp. 130-144.

CASE STUDY

Past work on adjacency graphs in architecture and urban planning is diverse and does not follow a common methodology. One exception is the work by White (1983; 1986), who captures the essence of functional planning using adjacency graphs as a driver in great detail. We have taken the example of an adjacency matrix printed in his book (see Figure 1) as starting point for occupation within this paper, and show how the different measures apply when this is transformed into a graph (note that an adjacency matrix is equivalent to an adjacency graph - every row of the [half-]matrix is transformed into a node of the same name; every colored entry into an edge between the rows at whose intersection it lies]).

Gephi

Having established the data, we needed to find a software with which architects and urban planners can immediately determine the sought graph measures. We firstly selected Gephi (Bastian et al. 2009 [1]; available from [2]), an open-source tool that is available for all platforms, for our initial test-drive. We noticed that there were two downsides:

1. The tool supports the entry of negative edge weights in principle, however, some algorithms and the visualization component do not take these into account. Thus, we left them away for the time being.
2. Furthermore, we also found hierarchical adjacencies ("interior" contains all interior spaces and "exterior" all exterior ones) in the graph which could not be accounted for. This seems to be commonplace, as further software packages (i.e. Pajek, which we have also look into) would not allow for that either. SNA, it seems, is not using hierarchies but rather concentrates on the 'flat case' - each space or function is deemed equal. Hierarchical adjacencies could, however, be constructed recursively, in the case that there were only relationships between the nested spaces or functions and the parent space or function would be a sufficient

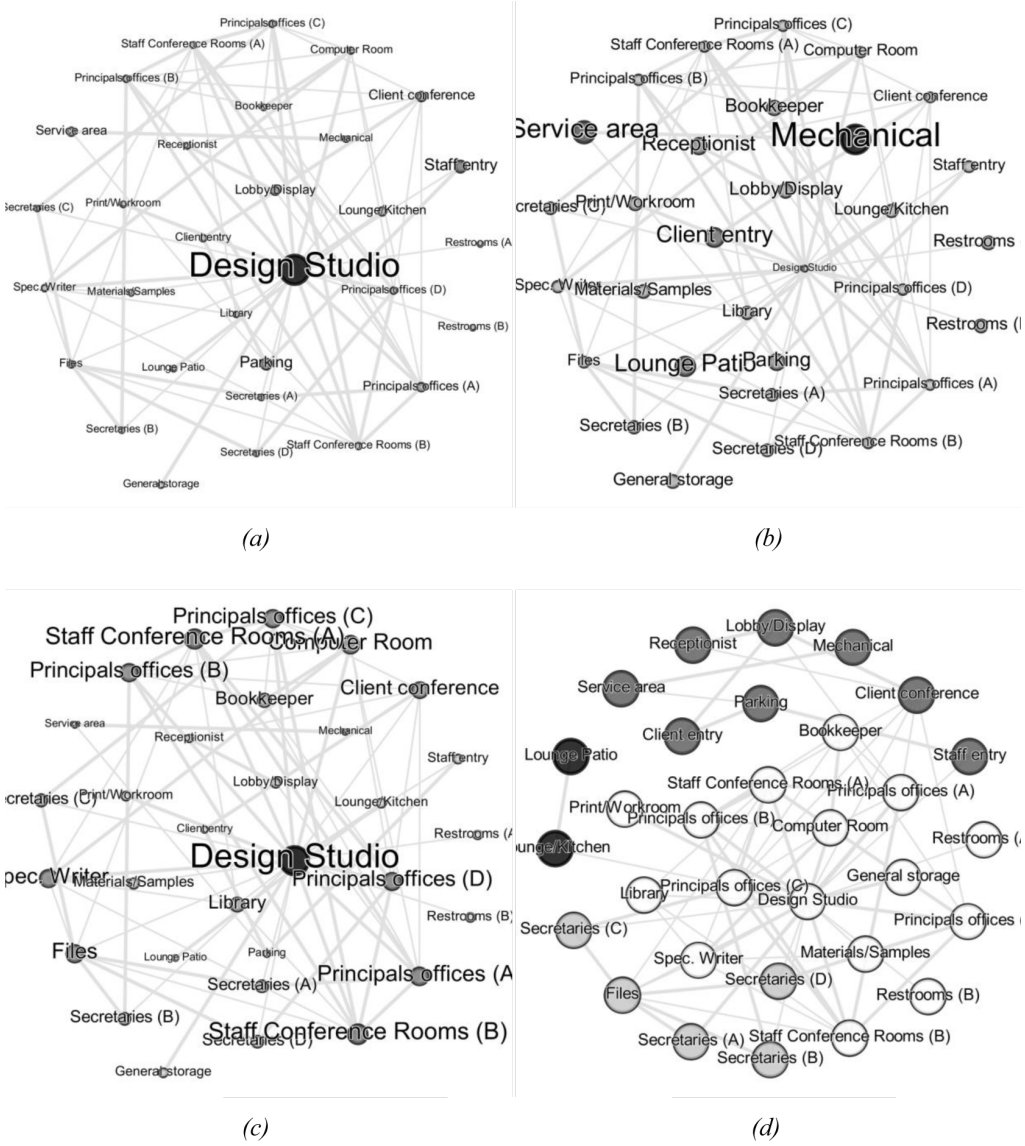


Figure 2
SNA measures in Gephi. (a) Betweenness, (b) Closeness, (c) Eigenvector centrality, (d) Cliques.

aggregation of all underlying spaces (the term "sufficient" demands that a space or function having sub-spaces/functions does in fact, on its level, represent all the underlying relations; if there are relations between different hierarchical levels, this will in general not hold).

Figure 2 shows the overall results of using Gephi, which we now want to examine in full detail:

Betweenness (Figure 2a) labels the 'Design Studio' as most central space. It is no wonder that this space is singled out, since it is the core function of the building and behaves as meeting space.

Closeness centrality (Figure 2b) is somewhat surprising: It labels 'Mechanical', 'Service area' and 'Lounge' as being closest to every other spaces, which might point to a planning error: Is 'Mechanical' really required to be a hub space with short distance to every other space? We think that this should not necessarily be the case. On the other hand, 'Design studio' (see middle of graph) seems to be quite dislocated (potentially long distances to other spaces).

Eigenvector centrality (Figure 2c) establishes an ordering in the relative importance of spaces, not from the standpoint of transitioning but from the standpoint of connectedness. In that view, 'Design Studio' (having the most degree!) would be the most important space, then come the offices and meeting/conference rooms. This is consistent with the overall function of the building.

Clique analysis (Figure 2d) uncovers four clusters of spaces, namely entrance areas (top part), core spaces (center), secretary spaces (lower-right part) and, interestingly, a kitchen/lounge area (right part).

From our study of Gephi, we concluded that the mentioned metrics can indeed be used to show (1.) different purposes of a building [e.g. through cliques as shown in Figure 2d] and (2.) uncover potential planning errors regarding adjacency without requiring any extra data [e.g. through analysis of connected components, not shown in Figure 2]. However, as the tool failed to include negative adjacencies, we clearly felt it as necessary to look at a different means of conducting our study. The choice fell onto Pajek (Batagelj and Mrvar 2003), which is commonly used because it is a software package that is free [but closed source!].

Pajek

Pajek itself was harder to handle as opposed to Gephi, since its visualization abilities are rather limited (mostly, researchers export graphs into a vector format and then work on beautify them in a different program) and there is need for a statistical/analytical background from the begin on since its user interface demands that one knows about the difference between a graph and the categories of measures taken on it (i.e. vectors on each node, clusters, permutations etc.). The centralities reported (see Figure 3) were partly in agreement with what Gephi had produced (Figure 3a: betweenness singles out Design Studio; Figure 3b - closeness, does not identify an overall top node as in the case of Gephi; Figure 3c: Eigenvector centralities do not exist per se; instead, there is a "hubs and authorities" model that seems to use Eigenvector centralities underneath, therefore sharing the same graph measures that are hard to perceive due to the lack in visualization capabilities of Pajek; Figure 3d shows Clique analysis, rather hard to perceive again [even though correctly labelling out Design Studio-Principals Office as clique, others being subordinate; certainly, the image could have been perfected by a proficient Pajek user using a different layout algorithm but this is what it really looks like for someone that is just trying to make the software 'show something interesting', in terms of adjacency analysis, without further insights what the statistical backgrounds were {and this is clearly refutable, to be sure! It is, nevertheless, the reality in terms of usability}].

DISCUSSION

There is no clear statement as to "thus, something needs to be different here" from seeing these two programs used for SNA. Certainly, Gephi excels at its visualization capabilities and Pajek does at its statistical options, which - as must be said - are far beyond

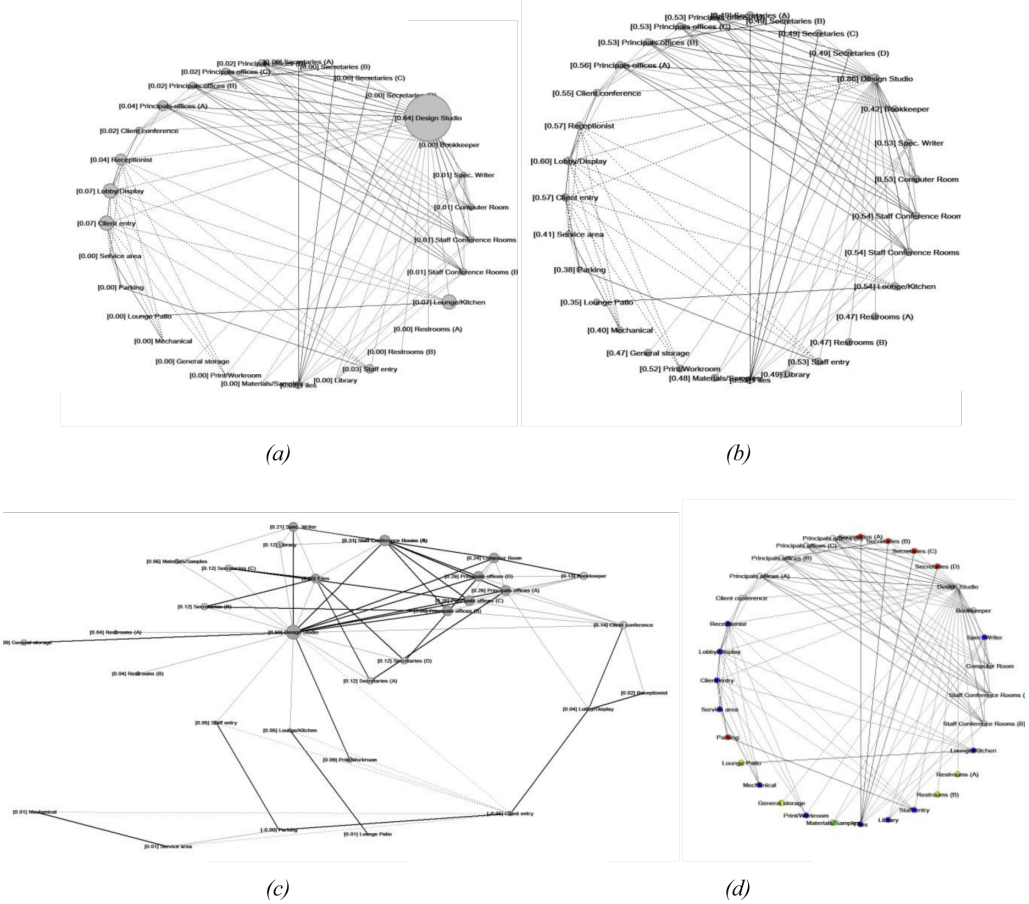


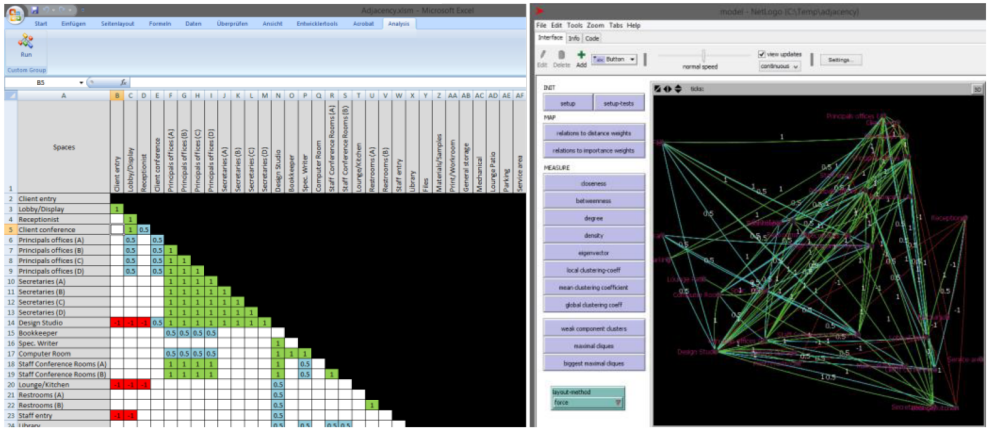
Figure 3
SNA measures in Pajek. (a) Betweenness, (b) Closeness, (c) Eigenvector centrality, (d) Cliques.

being comprehensible for an architectural office as it stands. What truly lies at the heart of the problem is that both are not accessible, in architectural terms. For this paper, we tried to make that crossing-over between both worlds more easily viable, using our own implementation which we coined "SpaceBook".

SPACEBOOK

SpaceBook is a mash-up between Excel and NetLogo, utilizing the ease in which it is possible to input adjacency relationships in Excel (Figure 4a) and the analytical capabilities of the NetLogo Network Extension. On a technical level, we use a self-written extension called *NetLogoExcelBridge* which allows NetLogo to be run from within Excel. Upon being started, NetLogo establishes a channel through which it can read and write data from Excel, which we then use to get adjacencies given as half-matrix (Figure 4a) and analyze/visualize that on a graph level (Figure 4b).

Figure 4
SNA in SpaceBook.
(a) Input of an adjacencies in Excel,
(b) computation of graph measures in NetLogo.



(a) *(b)*

Customizing graph analysis algorithms for adjacency relationships

Being open-source, NetLogo makes it easy to understand and extend the graph measures as well as the graph visualizations which are the key to the discussion we want to raise with this paper. As an example, we could easily re-implement closeness and eigenvector centrality so as to take negative relationships into account. In that process, we noticed that relationships need to be mapped to edge weights, since centrality algorithms have different semantics and thus expect different input (also see Table 1):

- For **closeness** and **betweenness**, the relations (negative=-1, desirable=0.5, mandatory=1) need to be mapped to positive weights since the underlying algorithm expects distances between nodes as input. If the relation is negative, we map this to a high number that simulates "infinity" (1000000 in our case). Desirable relations are assigned a positive weight (1), mandatory ones are mapped to 0 (i.e. there is "no cost" when crossing between the linked spaces or functions).
- **Eigenvector centrality** determines the im-

Table 1
Overview of examined graph measures, the necessary mapping from relations to weight as well as purpose for adjacency analysis.

	relation to weight	use for determining..
Closeness	(-1, 0.5, 1) → (∞, 1, 0)	closests/farthest nodes
Betweenness	(-1, 0.5, 1) → (∞, 1, 0)	hub nodes
Eigenvector	(-1, 0.5, 1) → (0, 1, 2)	important nodes
Degree		grade of specification
Density		of a node
Clustering coefficient (local, avg., global)	- (remove negative)	if adjacent nodes are connected as well
Clique analysis (maximal cliques)		clusters of highly inter-dependent nodes

portance of each node, based on the weight of the surrounding edges. Therefore, negative relations must map to zero (unimportant), all other should be positive (important, 1 and 2 in our case).

- **Degree centrality, density, clustering coefficient and clique analysis** do not use edge weights but examine the number of edges for a node. Thus, no mapping needs to be done in that case. However, we need to remove edges which have a negative relation in order to be able to compute these measures.

After performing the input mapping and tailoring the algorithms such that they can use negative weights and disregard isolated nodes without edges, we were able to repeat our experiments using the Design Studio data from White (1986; pp. 130-144). Since graph measures per se are mathematical tools not targeted at adjacency analysis in architectural terms, we tried to narrow down their specific purpose in the context of adjacency analysis (see again Table 1):

- Proximity and circulation are addressed by **closeness** and **betweenness centrality**.
- Importance of each space or function within the set of all functions can be examined using **eigenvector centrality**.
- Planners tend to attribute adjacencies only when they see them as being "important". Thus, it makes sense to consider how "well-specified" each node is in that regard, which can be done using either **degree** or **density centrality**.
- Spaces or functions tend to form groups. The two measures that can be used to measure that tendency are the **clustering coefficient** and **clique analysis**. Clique analysis reports subsets of the graph that are fully connected - i.e. spaces or functions which are highly interdependent, while the clustering coefficient tells us about the connectedness of the neighbors of a node (in SNA terms, this would be: "Do my friends know each other, too?").

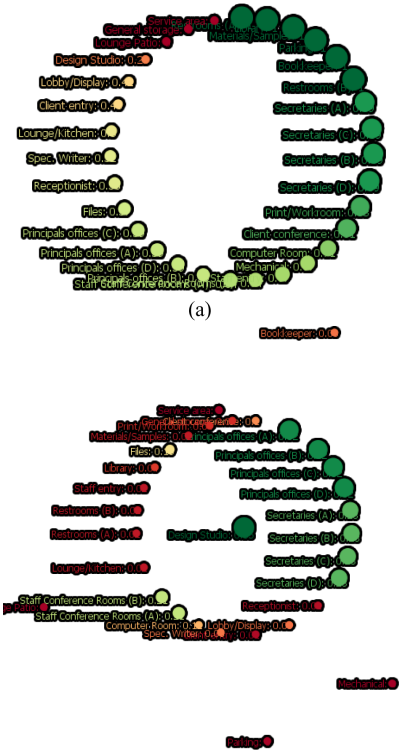


Figure 5
Visualizations in SpaceBook. (a) Circular layout, (b) radial layout showing hierarchies, (c) cluster layout. Original sceenshots from within SpaceBook - black background has been turned into transparent in order to conserve some ink.

Customizing the visualization of results

In general SNA software, the layout is applied in a post-step. However, as this approach specifically deals with adjacency analysis, we thought it reasonable to merge the computation and visualization into one step, since we already know what we are looking for when calling up the computation of a graph measure. Netlogo already offers some layouting options (e.g. layout-spring based on Fruchterman and Reingold 1991), and these can be extended either in the back-end (Java) or in the NetLogo language itself (which is what we did). We used three types of layouts (1.) circular layout (Figure 5a) in which each node's importance is also encoded into node size and color, a hierarchical layout which is radially centered around the most prominent node (Figure 5b) and a custom-written clustering layout, which lays out all supplied clusters supplied by clique analysis in a grid (Figure 5c). The latter was a bit more complex, since a node can belong to many clusters. Merging computation and visualization into one step is a good approach for exploring the data. The results can also be written in tabular form into Excel for further processing.

CONCLUSION AND OUTLOOK

This paper has presented a case study which uses graph measures from Social Network Analysis in order to analyze adjacency relations. From examining two software packages (Gephi, Pajek) on adjacencies of a design studio taken from literature, two main observations were made: (1.) There is a slight variation in the computation of the graph measures used, owing to different semantics and the fact that negative relationships (commonplace in adjacency analysis) might not be taken into account. An added issue in that context is that the underlying algorithms cannot always be scrutinized, since such "end-user" packages may not be open-source. (2.) The same goes for the graph layout/visualization features, which may additionally be too general to be useful for the problem domain at hand.

In order to address both of these problems, we have conceived SpaceBook, a mash-up between Excel and Netlogo that is able to analyze and visualize adjacency in a meaningful way, with the ease of a click in Excel. We are positive that this approach is well-integrated into the workflow of early-stage design, which requires rapid turnaround between data entry and analysis. As a difference to classical Social Network Analysis, we want to note that the calculated measures need to be interpreted qualitatively rather than quantitatively (e.g. for indication of relative rank instead of 'exact' centralities for each node), since data at that stage is still conceptual. It can nevertheless be nswer question such as "are there any ill-specified spaces with no relationships" and "what are the most prominent nodes in the adjacency network". As outlook, we want to conduct the same case study with a much larger adjacency graph such as in the hospital domain, which seems interesting because there are also hierarchical relationships (subgraphs contained in a node).

ACKNOWLEDGEMENTS

The authors wish to acknowledge the valuable insights into SNA given by FASresearch (especially Harald Katzmair and Christian Gulas). Furthermore, we wish to acknowledge the important role of the Center for Computationally Complex Systems at TU Wien in bringing disciplines together (especially Niki Popper and Felix Breitenecker), as well as Institute for Managing Sustainability (especially Markus Hametner, Asya Dimitrova and André Martinuzzi) for bringing us into the field of SNA in project terms.

REFERENCES

Bastian, M, Heymann, S and Jacomy, M 2009 'Gephi: an open source software for exploring and manipulating networks', *International AAAI Conference on Weblogs and Social Media*

Batagelj, V and Mrvar, A 2003, 'Pajek - Analysis and Visualization of Large Networks', in Jünger, M and Mutzel, P (eds) 2003, *Graph Drawing Software*, Springer, Berlin, pp. 77-103

Bonacich, P 1972, 'Techniques for analyzing overlapping memberships', *Sociological Methodology*, 4, pp. 176-

185
Bonacich, P 2007, 'Some unique properties of eigenvector centrality', *Social Networks*, 29, pp. 555-564
Borgatti, S, Mehra, A, Brass, D and Labianca, G 2009, 'Network Analysis in the Social Sciences', *Science*, 323, pp. 892-895
Fidler, D and Hanna, S 2015 'Introducing random walk measures to space syntax', *Proceedings of the 10th Space Syntax Symposium*, pp. 145:1-145:9
Freeman, LC 1977, 'A set of measures of centrality based upon betweenness', *Sociometry*, 40, pp. 35-41
Fruchterman, TMJ and Reingold, EM 1991, 'Graph Drawing by Force-Directed Placement', *Software - Practice and Experience*, 21(11), pp. 1129-1164
Hillier, B and Hanson, J 1984, *The Social Logic of Space*, Cambridge University Press
Holland, PW and Leinhardt, S 1971, 'Transitivity in structural models of small groups', *Comparative Group Studies*, 2, pp. 107-124
Law, S, Chiaradia, A and Schwander, C 2012 ''Towards a multi-modal space syntax analysis. A case study of the London street and underground network', *Proceedings of the 8th Space Syntax Symposium*, pp. 8021:1-8021:20
Noh, JD and Rieger, H 2004, 'Random Walks on Complex Networks', *Physical Review Letters*, 92(11), pp. 118701:1-118701:4
Porta, S, Crucitti, P and Latora, V 2006, 'The Network Analysis of Urban Streets: A Primal Approach', *Environment and Planning B*, 33(5), pp. 705-725
Stephenson, K and Zelen, M 1989, 'Rethinking centrality: methods and examples', *Social Networks*, 11, pp. 1-37
Wang, W-C and Liao, H-J 2007 'Implementing space syntax in an open-source GIS: GRASS GIS approach', *Proceedings of the 6th Space Syntax Symposium*, pp. 100:01-100:14
Watts, DJ and Strogatz, S 1998, 'Collective dynamics of "small-world" networks', *Nature*, 393(6684), p. 440–442
White, ET 1983, *Site Analysis*, Architectural Media
White, ET 1986, *Space Adjacency Analysis*, Architectural Media
Wilensky, U 1999, *NetLogo (Software), ccl.northwestern.edu/netlogo/*, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL
[1] https://gephi.org/publications/gephi-bastian-feb09.pdf
[2] https://gephi.org/
[3] http://www.iemar.tuwien.ac.at/processviz/spacebook