Salvini, M. M., Gnos A. U. and Fabrikant, S. I. (2011): COGNITIVELY PLAUSIBLE SPATIALIZATION OF NETWORK DATA. *Proceedings, 25th Cartographic Conference*, International Cartographic Association, Jul. 3-8, Paris, France

COGNITIVELY PLAUSIBLE SPATIALIZATION OF NETWORK DATA

SALVINI M.M., GNOS A.U., FABRIKANT S.I.

University of Zurich, ZURICH, SWITZERLAND

We present a network visualization approach based on the spatialization framework proposed by Fabrikant and Skupin (2005) which allows the perceptually salient and cognitively plausible information visualization of large relational datasets at high aggregation level. The framework is put to a rigorous test in a case study aimed at uncovering the latent relational structure of global cities, based on air passenger volumes. A force-directed placement algorithm is employed to depict cities (i.e., nodes) and respective passenger flows between the cities (i.e., edges) as a network in relational data space. The relational dataset is further semantically and cartographically generalized by means of common network clustering methods coupled with the adaptive multiplicative weighted Voronoi algorithm (AMWVD). Qualitative expert interviews have been carried out for evaluating the resulting generalized network visualizations. The evaluation confirms the overall validity of methodological framework and suggests further steps for refinement.

INTRODUCTION

Vast amounts of relational data are becoming available through easily accessible online databases such as, Facebook, Wikipedia, and the WWW, to study the networked information society. Effective and efficient methods for the analysis and visualization of such large relational datasets have gained in importance in the last years for various research communities, mostly outside of cartography. The multivariate complexity of information networks poses specific challenges for visual information exploration, which opens the door for novel and alternative visualization methods (Viégas and Donath 2004).

One problem in network analysis and visualization today is that commonly accepted methods have been introduced at a time when the researched networks consisted of typically only a few dozen nodes and links (Viégas and Donath 2004). Networks analyzed today may have tens of thousands nodes and links, and thus the classic network visualization approaches break down, as can be seen in Figure 1, where in fact no structure can be discovered at all! Commonly applied graph drawing methods, typically based on graph aesthetics measures such as, minimizing crossing edges, symmetry, etc. (Di Battista et al. 1994) have rarely, if not at all, been empirical evaluated (Ware et al. 2002). One problem with graph aesthetics measures is that they lack solid theoretical and empirical foundations. It is still not clear today what aesthetics criterion (or which combinations) and for which particular use context can improve a network layout, as desired by the graph drawing community (Purchase 1998, Bennett et al 2007).

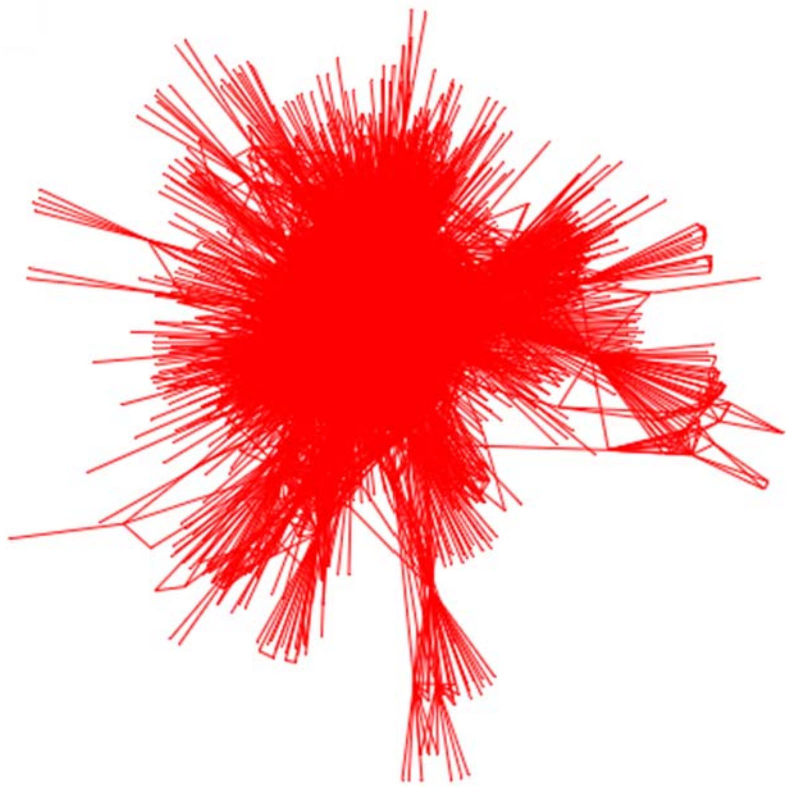

Figure 1: Global cities network spatialization based on air passenger flows 2004 (ENAC, 2004; IGUL 2008). Layout is generated with a Kamada-Kawai algorithm.

One main challenge in network visualization is to reduce data complexity by projecting a multivariate (origin-destination) data matrix onto a lower dimensional (i.e., two-dimensional) planar display space. This spatialization is achieved by selecting the thematically most important and perceptually most salient links and nodes without losing the overall structure of the original network. Dealing with multivariate data reduction and effective visuo-spatial representation in a two-dimensional plane is obviously one of the main goals of cartography. The applicability of cartographic and geovisualization methods for spatialization has been empirically validated in earlier work (Skupin and Fabrikant 2003; Fabrikant et al. 2004), and the benefit of applying these methods to non-spatial data is well documented (Couclelis 2002; Fabrikant and Skupin 2005). Other visualization communities and fields thus may well benefit from the rich body of work available in cartography (Slocum et al. 2010).

In this paper, we present a perceptually salient and cognitively plausible information visualization approach for large relational datasets. This approach is based on the spatialization framework proposed by Fabrikant and Skupin (2005). The framework is put to a rigorous test in a case study aimed at uncovering the latent relational structure of global cities, based on air passenger volumes (ENAC, 2004; IGUL 2008). The connectivity of the cities is expressed with the number of airline passenger between them. This is an established method in economic geography to approximate the relationship between cities in studies about global city networks. However, this approach also includes some critical limitations, for example, the occurrence of touristically attractive destinations being over represented in airline travel data (Derudder and Wiltox 2008). We are aware of this data limitation in our case study, but this has no implication for our methodological contribution.

SPATIALIZATION FRAMEWORK

The spatialization framework (Fabrikant and Skupin 2005) suggests "*the systematic transformation of high-dimensional data sets into lower-dimensional, spatial representations for facilitating data exploration and knowledge construction*" (Skupin and Fabrikant 2007). It also allows the perceptually salient and cognitively plausible information visualization of large relational datasets. Similarly to cartographic generalization, the spatialization framework includes semantic and geometric generalization.

The semantic generalization process in spatialization relates to the identification of the appropriate spatial metaphors which captures the essential characteristics of the data entities to be visualized (Fabrikant and Skupin 2005). Sound metaphors not only combine semantic properties from a source domain, but also ideally contain cognitive and experiential aspects (Fabrikant and Skupin 2005). Examples of such metaphors are the landscape metaphor for a continuous data space, the city metaphor for discrete data spaces, the scale metaphor for underlining the change in level of detail, among others. Fabrikant and Skupin (2005) identify four semantic primitives as the building blocks of more complex spatial metaphors. These primitives are applicable to a range of information types, and are associated with a range of geographic source domains. The four semantic primitives are: locus, trajectory, boundary, and aggregate (Fabrikant and Skupin 2005). *Locus* is characterized by a location a two-dimensional representational space which is determined by its semantic relationships with other information items in this space. *Trajectory* is a linear entity type, which underlines the relationship between items. *Boundary* is also a linear type of representation and it captures discontinuities in an information space. Boundaries delineate semantic regions. Semantic regions are called aggregates, and they represent an areal entity type. An *aggregate* is the result of a classification process; it is understood as a homogenous zone (with or without a discrete boundary) that can be distinguished from other zones (Fabrikant and Skupin 2005).

The geometric generalization deals with the perceptually salient depiction of the semantic primitives. As soon the semantic primitives are assigned, they can be straightforwardly represented graphically, using Bertin's (1967) commonly known visual variables (Fabrikant and Skupin 2005). For example, depending on the displayed scale, the semantic primitive locus may be represented as a point or an area, the linear primitive trajectory and boundary by a line, and the aggregate primitive by a point or a polygon (Fabrikant and Skupin 2005).

SEMANTICS OF THE GLOBAL CITY NETWORK

Different spatial metaphors capture different characteristics of the global city network. We identify two main metaphors: the distance-relatedness, and the scale metaphor.

The distance-relatedness metaphor is actually a modification of the well-known distance-similarity metaphor (Fabrikant et al. 2004), and is also based on the commonly known *"first law of geography: everything is related to everything else, but near things are more related then distant things"* (Tobler 1970). In other words, near items (cities) in the topological space are more strongly connected than distant cities.

The semantic primitive locus captures the placement of cities, represented as nodes, in our network data space. City (i.e., node) placement is dependent on a city's interactions with other cities. In traditional network analysis and visualization, edges are used to emphasize the relations between the nodes. In our approach, we also emphasize the discontinuity between nodes with the boundary metaphor, to show cities that are not at all or only weakly connected.

The second metaphor used in our example is the scale metaphor. The scale metaphor is not only fundamental to geographical analysis, but it is also associated with cognitive and experiential properties of the real world (Fabrikant 2001a), such as human perception and cognition of geographical phenomena and processes. As we change the viewing scale, the thematic focus changes, for example, from the single field to a complete landscape. The scale metaphor allows us to solve the data density problem in large networks by establishing a hierarchical order of data items based on different levels of detail. Similarly to topographic map series, on a smaller scale, the focus of interest may be on regions or larger cities, while at larger map scale, smaller cities and towns become more relevant.

DATA TRANSFORMATION AND GEOMETRIC IMPLEMENTATION

For the geometric generalization step we propose an interdisciplinary approach which integrates methods from geovisual analytics, cartographic design, social network analysis (Wasserman and Faust 2008), and information visualization, including graph drawing (Di Battista et al. 1994).

As described earlier, we are interested in highlighting a city's location in a topological information space, based on air passenger data, by means of the distance-relatedness metaphor. Graph drawing algorithms (Di Battista et al. 1994) seem useful for this step. With *DrL* (Davidson et al. 2001), we chose a force directed placement algorithm to transform the data matrix into a network (Chen 2004). The force directed placement algorithm places strongly connected cities closer together on the network than less connected ones.

Next, we are interested in depicting the underlying semantic hierarchy in the city network. At smaller (coarser) scale we want to highlight regions and larger cities. Both, city size and regional organization can be defined in a cities network. Highly connected cities form city communities or urban regions. These regions are defined by high within-community connectivity, and sparse relations between other communities (Newman and Girvan 2004). We calculate city communities with the Newman and Girvan (2004) algorithm which iteratively removes links between cities with the greatest betweenness. Betweenness is defined as the sum of all shortest connections between every single node in the network (Wasserman and Faust 2008).

This coarser level of the semantic hierarchy is depicted with homogeneously colored community zones, to visually emphasize the containment principle, as one key characteristics of the hierarchical organization (Fabrikant 2001b). City regions on smaller scale vary in size proportionally to passenger flow volumes between cities. We employed the adaptive multiplicative weighted Voronoi method (AMWVD) to achieve this (Reitsma et al. 2007). The AMWVD is in essence an extension of the classical Verona tessellation. We implemented an OpenJump GIS plug-in, which transforms given nodes into AMWVD polygons, and additionally creates a convex hull that includes all generating nodes. As the AMWVD method is computationally expensive, we calculated the AMWVD only for the most thematically salient cities in the network, which is sufficient for our case study. The saliency of a city is given by the quantity and quality of its relations to other cities. Guimerà and Amaral (2005) developed a detailed taxonomy of nodes based on their functions in a network. For example, the most important cities in a network are called hubs. Hubs have a high degree of network centrality, and are defined by many within- and between-community connections. The between-community connections are typically with other hubs. The result of the AMWVD spatialization is shown in Figure 2 below.

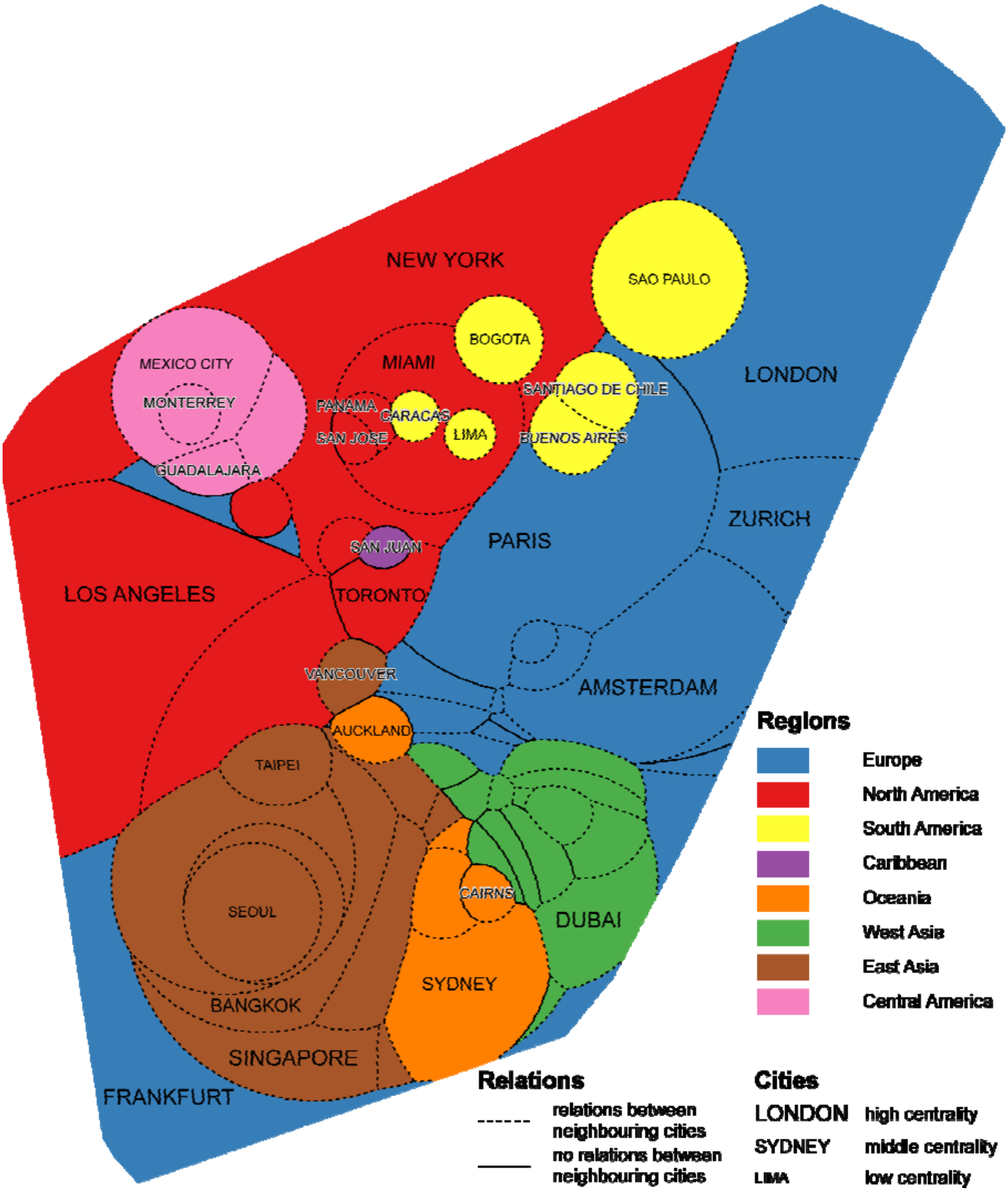

Figure 2: Global cities network based on air passenger flows 2004 at the regional level.

The cities shown in the AMWVD in Figure 2 are all considered hubs in the passenger flow data space (ENAC, 2004; IGUL 2008). The AMWVD reveals hub polygons scaled according to annual air passenger volumes. The hub labels are placed within the largest polygon fragment (explained further down) of the AMWVD polygons, to indicate the relative location of the hub to other hubs. Additionally, label size indicates the degree of node centrality. More central hubs have larger labels than less central ones. For example, the growing importance of Dubai as a recent hub in international air traffic becomes apparent in this way.

We use the visual variable color hue to distinguish the community membership of cities. We employed ColorBrewer2 to assign cartographically sound colors emphasizing the qualitative aspect of our data (Brewer et al. 2003). The hub regions (e.g., the East Asian hub cluster in brown) emerge as a result of the within/between connectivity analysis described earlier. The uncovered regions appear meaningful at first glance, as hubs within the same geographic region appear near each other in the spatialization. For example, Zurich, London, Paris and Amsterdam are all located within the blue European cluster. On the one hand, this seems unsurprising, as airplane movement is typically constrained by a finite flight range, thus the geographical configuration is implicitly re-produced in the spatialization. However, we might also discover a cultural effect in Figure 2. For example, Auckland (NZ) is placed not only close to East Asian cities (logical in terms of flight distance), but also to linguistically similar North American hubs, and culturally close western European cities.

The semantic primitive *boundary* is also used in Figure 2 to depict the kind of relationship between the different cities (i.e., hubs) in the global city network. The visualization of the boundary primitive is perceptually necessary, as the AMWVD might create polygons that share a common border, but that are not necessarily directly related with each other. We employ dashed boundary lines to visually distinguish adjacent polygons which have direct network connections (i.e., are permeable via hubs), from those that do not. Zones that are not connected directly via hubs are distinguished with a solid boundary line. This literally means that no direct passenger flow exist between these polygons. For example, the adjacent cities Zurich and London are directly related, as they share a permeable (dashed) polygon boundary, whereas London and Paris surprisingly seem not related, as they are separated by a solid boundary. It so happens that the flight connections between Paris and London were not included in the used dataset, and this is revealed by the spatialization.

Figure 2 also exemplifies a more general issue with the AMWVD method, as fragmented (i.e., discontinuous) polygons can occur as mentioned earlier (i.e., Frankfurt). The polygon labeled Frankfurt at the lower left corner of Figure 2 belongs to a much larger, but fragmented polygon for this city. The generator point for Frankfurt lies within an unlabeled blue zone to the East of Auckland, with unlabeled green polygons at its southern border. This can happen when generator points that are close to each other also have higher weights than the other nodes in the database. Large polygons bend beyond the spatial extent of the convex hull, and reenter the map where there is enough space left for expansion, and thus might enclose other smaller polygons as a result (Reitsma et al. 2007).

EVALUATION

Qualitative expert interviews (including data experts and visualization experts) have been carried out for evaluating the resulting network visualizations at various levels of detail. For this paper, we report only on the results for the region level (Figure 2). For most of the experts the visualization was completely unfamiliar. This can of course influence its evaluation, as we did not explain the displays at the outset, and did not give experts respective training for correct interpretation. Overall, the evaluation confirms the validity of a cartographically sound methodological framework. The experts find the network visualizations thematically meaningful, and confirm that the chosen visualization approach indeed emphasizes cluster structure in the data: cluster membership of a city is easily visible and understandable through meaningful relative location and cluster assignment by color hue. The added benefit of simplification through aggregation (i.e., points to polygons) is somewhat reduced by the general unfamiliarity with this kind of visualization, and respective resulting of unintuitive geometry (i.e., island polygons). All experts found the interpretation of the island polygons (as one typical artifact of the AMWVD) conceptually difficult to understand. Moreover, island polygons were also found to be perceptually problematic, as the assessment of disconnected polygons makes magnitude judgment and comparison more difficult. Interestingly, in an empirical study by Reitsma and Trubin (2007), where participants were specifically asked to estimate polygon sizes in two standard continuous tessellations, compared to the AMWVD, the researchers did not find any evidence that size estimation with the AMWVD was more difficult or included more errors, compared to the two other continuous tessellation methods. Clearly, the careful application of Bertin's visual variables also helps to improve the readability of the visualizations.

SUMMARY AND OUTLOOK

In this paper, we detail our network spatialization approach based on a theoretically sound spatialization framework, coupled with systematic use of cartographic depiction methods. We apply the proposed framework to identify a global city network, based on global air passenger flow data. The distance-relatedness metaphor and the scale metaphor capture the main semantic characteristics of this network. For the geometric generalization step we propose a novel approach which integrates a forced directed placement algorithm (DrL) with adaptive multiplicative weighted Verona (AMWVD) polygons, for the thematically relevant and perceptually salient visualization of networks at coarser levels of detail. A qualitative expert evaluation confirms the overall validity of this network visualization approach based on simplification through aggregation (i.e. points to polygon transformation). However, our empirical results also raise new questions about the cognitive adequacy of the AMWVD approach, especially relating to the complex, and somewhat unintuitive geometry. A future research avenue therefore will be to further explore the potential of the AMWVD in the network context. On the one hand, the ambivalent evaluation outcome could be due to inherent limitations of the algorithm, which would then require algorithm redesign, or it might be simply related to user unfamiliarity, which—like map reading in general—could be solved with adequate training, to yield the expected benefits.

ACKNOWLEDGMENTS

We would like to thank our experts who provided valuable information by serving as research participants in the interviews. We are also grateful to René Reitsma who introduced us to his intriguing AMWVD ideas during his sabbatical stay at our research group in early 2009.

REFERENCES

Bennett, C.; Ryall, J.; Spalteholz; and Gooch, A. 2007. The Aesthetics of Graph Visualization. *Computational Aesthetics in Graphics, Visualization, and Imaging* **:1-8.**

Bertin, J.; and Barbut, M. 1967. *Sémiologie graphique: les diagrammes, les résaux, les cartes.* Mouton, Paris.

Brewer, Cynthia A., Geoffrey W. Hatchard and Mark A. Harrower. 2003. ColorBrewer in Print: A Catalog of Color Schemes for Maps, *Cartography and Geographic Information Science* **30(1): 5-32.**

Chen, C. 2004. *Information Visualization. Beyond the horizon*.

Couclelis, H. 1998. Worlds of information: The geographic metaphor in the visualization of complex information. *Cartography and Geographic Information Systems* **25(4): 209-220**.

Davidson, G. S., Wylie, B. N., & Boyack, K. W. (2001). Cluster stability and the use of noise in interpretation of clustering. *Proceedings,* IEEE Information Visualization (Infovis 2001), San Diego, CA**: 23-30.**

Derudder, B.; and Witlox F. 2008: Mapping world city networks through airline flows: context, relevance, and problems. *Journal of Transport Geography* **16: 305–312**.

Di Battista, G.; P. Eades; R. Tamassia; and I.G. Tollis. 1994. Algorithms for drawing graphs: an annotated bibliography. *Computational geometry* **4: 235-282**.

Fabrikant, S.I. (2001a): Evaluating the Usability of the Scale Metaphor for Querying Semantic Space. *Lecture Notes in Computer Science* **2205.**

Fabrikant, S.I. (2001b): Visualizing region and scale in information space. *Proceedings,* The 20th International Cartograpic Conference, ICC 2001, Beijing, China***:*** **2522-2529.**

Fabrikant, S.I. and A. Skupin. 2005. Cognitively Plausible Information Visualization. *Exploring Geovisualization*, ed. A.M. Dykes; A.M. MacEachren; and M.J. Kraak.

Fabrikant, S.I.; Montello, D.R.; Ruocco, M.; Middleton, S. 2004. The Distance-Similarity Metaphor in Network-Display Spatialzations. *Cartography and Geographic Information Science* **31 (4): 237-252.**

Garlandini, S., Fabrikant, S.I. (2009): Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. *COSIT 2009*. Springer, Berlin, **195-211**.

Guimerà, R. and L.A.N. Amaral. 2005. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics* **2: 1-13**.

Newman, M.E.J. and M. Girvan. 2004. Finding and evaluating community structure in networks. *The physical review E*. **69 (2): 26113.**

Purchase, H.C. 1998. Performance of Layout Algorithms: Comprehension, not Computation. Journal of Visual Languages and Computing **9 (6): 647-657**.

Reitsma, R. and S. Trubin. 2007. Partioning with adaptive Voronoi diagrams. *Information Visualisation* **6: 123-138.**

Reitsma, R., S. Trubin; and E. Mortensen. 2007. Weigth-proportional space partioning using adaptive Voronoi Diagrams. *Geoinformatica* **11: 383-405**.

Skupin, A. and S.I. Fabrikant. 2007. Spatialization. In *The handbook of geographical information science*, ed. J.P. Wilson and A.S. Fotheringham, **61-79**.

Skupin A. and Fabrikant, S.I. 2003. Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. *Cartography and Geographic Information Science,* **30 (2): 95-119.**

Slocum T.A.; McMaster, R.B.; Kessler, F.C.; and Howard, H.H. 2009. *Thematic Cartography and Geovisualization – Third Edition*. Pearson, Upper Saddle River (USA).

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46(2): 234-40.**

Viégas, F. and J. Donath. 2004. Social network visualization: can we go beyond the graph? In *Social Networks for Design and Analysis: Using Network Information in CSCW*.

Ware, P.; Purchase H.; Colpoys, L.; and McGill, M. 2002. Cognitive measurements of graph aesthetics. *Information Visualization* **1(2): 103-110**.

Wasserman, S.; and K. Faust. 2008. Social network analysis: Methods and applications. New York.