Visualizing Global Patterns in Huge Digraphs

Noel Patson

School of Computing Sciences

Faculty of Arts, Business, Informatics and Education, CQUniversity

Rockhampton, Australia

Email: n.patson@cqu.edu.au

Abstract—**Traditional methods for depicting directed graphs are inadequate when the number of vertices and edges are too large. Also underlying geographical coordinates are generally not considered. A new visualisation method is presented which transforms graphs into a 3D surface which can be shown as a contour map superimposed over a geographical region thereby preserving the underlying geographical details.**

Keywords-**information visualization; digraphs;**

I. Introduction

Huge digraphs arise in areas such telecommunications, computer networks, the internet, banking, commerce, ecology and social networks, etc.. In many of these areas the location of the digraph nodes is significant in understanding the overall graph from a geographical perspective. The traditional method for visualizing directed graphs, arrows connecting dots, is hampered with obfuscation whenever the number of nodes and edges is large. Also the traditional digraph data structure, the adjacency matrix, does not integrate the geographical location of nodes. Usually, visualization approaches for depicting large digraphs have sought to reposition the nodes so as to minimize overlapping and thus detach themselves from the underlying geographical setting [1]–[11]. In this paper a new visualization method is presented in which the underlying geographical context of the digraph is important. Emphasis is made on revealing large scale structures rather than fine details. If the metaphor is used of associating the visualization of digraphs to the observation of clouds then the traditional method is akin to viewing cloud water droplets from within the cloud using a magnifying glass while the presented method is like viewing the clouds at a distance whereby cloud formations can be discerned. Detailed digraph information is sacrificed in order to understand global patterns.

The visualization focus is on exploring the overall input or output of nodes set in the digraph's geographical context. The main thrust of the method is to convert the discrete information of digraphs into a smooth 3D surface which can be shown as contours overlaid with geographical identifiers such as boundaries and city names. The method has been granted a PCT patent [13].

II. Turning Digraphs into Weather Maps

To illustrate the method consider a simple weighted digraph with 3 nodes and 6 edges as shown in the traditional way in Figure 1(a) and as an adjacency matrix in Figure 1(b). As a first step in translating the digraph into the node output form of digraph "weather map" the edges are converted into vectors directed out of each node as shown in Figure 1(c). The length of each vector is proportional to the weight of the associated edge. In the node input form, the same procedure is applied except that the direction of the edges is considered to be in the reverse direction.

A. Turning Vectors into a Surface

In the next stage the vectors emanating from each node are translated outward from the node so that the beginning of each vector is set on a circle centered on the node as illustrated by the 3 vectors and single node in Figures 2(a) and 2(b).

The radius is dependent on a parameter κ which used in the surface definition formula. The value $R_\kappa = 1/\sqrt{\kappa}$ is chosen so that surface peaks aggregate over digraph nodes.

B. Definition of Surface

Refer to Figure 3 for an illustration of the terms used in this definition. Consider $\{\underset{\sim}{\nu}_r; r = 1, 2, \cdots, n\}$ to be the set of n translated vectors that represent digraph edges. For each $r = 1, 2, \cdots, n$ define $\underset{\sim}{\varphi}_r$ to be the vector obtained by the rotation of $\underset{\sim}{\nu}_r$ through an angle $\pi/2$ anticlockwise. Let $\underset{\sim}{m}_r$, $r = 1, 2, \cdots, n$ be the position vectors pointing to each $\underset{\sim}{\varphi}_r$. Associated with every point (x, y) in $\mathbb{R}^2$ there is a set of n position vectors such that for $r = 1, 2, \cdots, n$:

$$\underset{\sim}{\omega}_r(x, y) = (m_{r1} - x)\underset{\sim}{i} + (m_{r2} - y)\underset{\sim}{j} + 0\underset{\sim}{k}.$$

IEEE
computer
society

(a) Simple digraph. (b) Adjacency matrix. (c) Digraph represented as a set of vectors
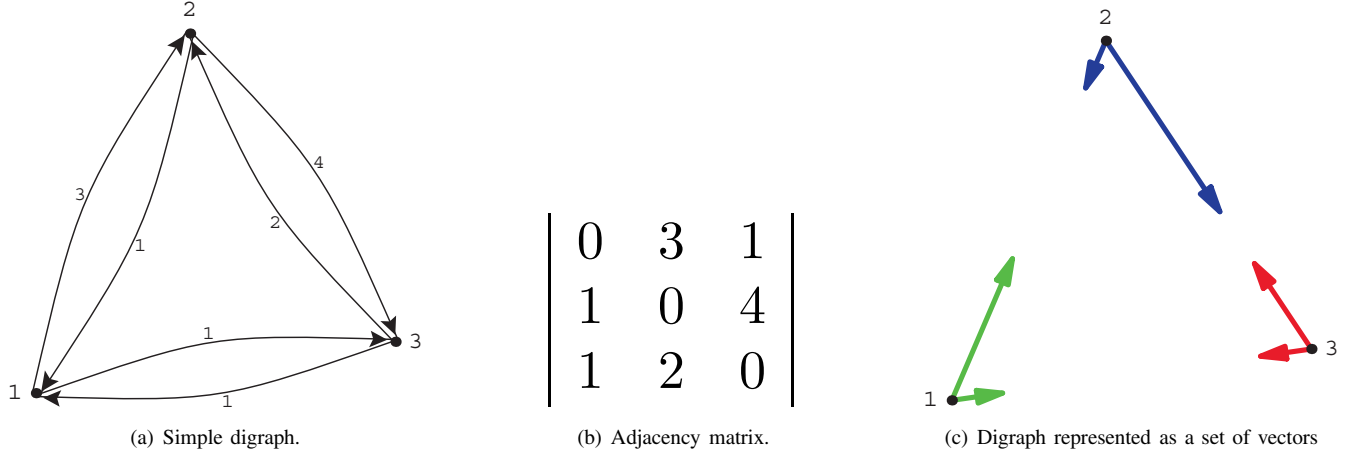
Figure 1.   A simple digraph represented in 3 different ways [14], [15].

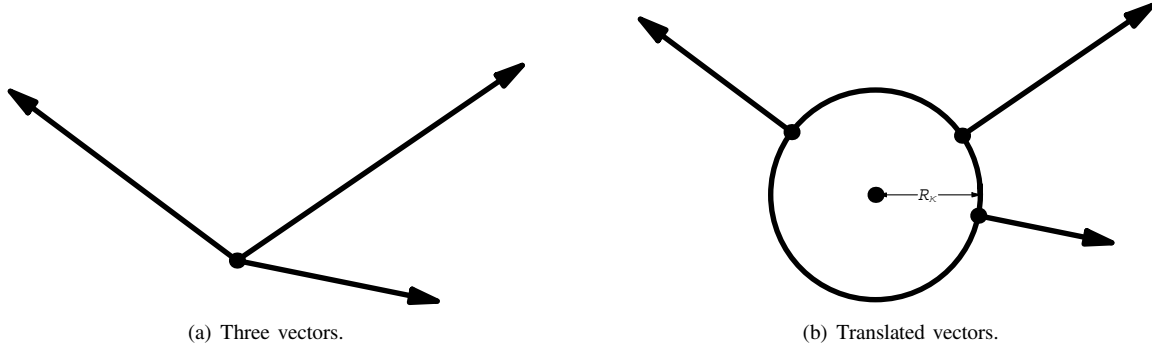(a) Three vectors. (b) Translated vectors.

Figure 2.   An example of three vectors from a vector representation of a digraph node (a) translated onto a circle centered on the node with a radius R_κ (b) [15].

The vector $\omega_r(x,y)$ points from (x,y) to m_r. Define vector μ at (x,y) as

$$\mu(x,y) = \sum_{r=1}^{n}(\omega_r(x,y) \times \varphi_r)e^{-\kappa|\omega_r(x,y)|^2}. \qquad (1)$$

where the magnitude $|\omega_r(x,y)|$, is the distance between the point (x,y) and the point m_r at the end of m_r.

A three dimensional surface results by applying the function 1 to the set of translated vectors derived from a digraph. The surface may be represented as a contour map so that geographical points of interest such as node locations can be identified. Figures 4(a), 4(b) and 4(c) show three variations of the surface arising from the digraph in Figure 1(a). In Figure 4(a) the value of κ results in smaller surface peaks and nadirs (valleys), surrounding nodes. A smaller value of κ results in larger surface deformations as can be seen in Figure 4(b). This is like "zooming" into the digraph. In Figure 4(c) only nadirs, that is the negative values of $\mu(x,y)$, are shown.

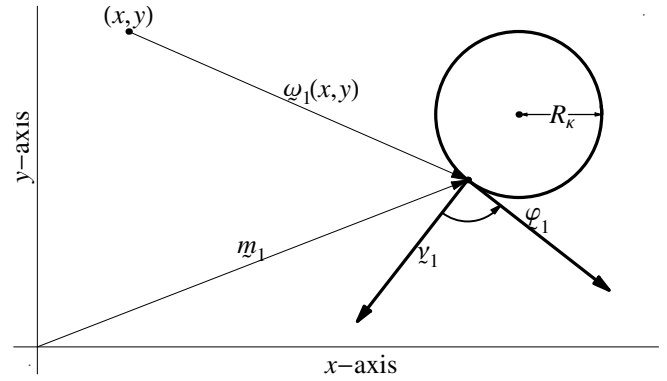The height of the peak which is located above each of the

Figure 3.   An illustration of the terms that are used in Definition II-B that determine a 3 dimensional surface [15].

nodes in Figures 4(a) and Figure 4(b) is proportional to the total weights of the edges going out from the corresponding node. The depth of each nadir surrounding each node represented in Figures 4(a), 4(b) and 4(c) is proportional to the weight of the edge that corresponds to the nadir [15].
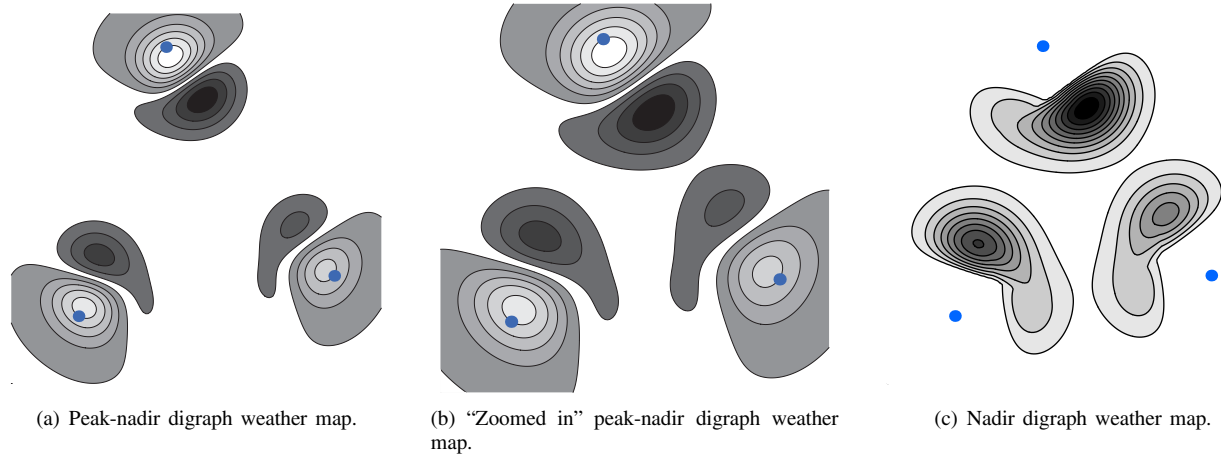
481

(a) Peak-nadir digraph weather map.

(b) "Zoomed in" peak-nadir digraph weather map.

(c) Nadir digraph weather map.

Figure 4.   Three weather map representations of the digraph in Figure 1 [15].

(a) A resultant vector representation of the digraph data. The vectors are pointing away from the points m_i which represent callers.

(b) With parameter $\kappa = 0.5$, the resolution is very fine showing details at a local level.

(c) With parameter $\kappa = 0.3$ the resolution shows less detail and highlights connections at a higher level.

(d) With parameter $\kappa = 0.1$ the resolution is very coarse, so that large scale global features are exposed.
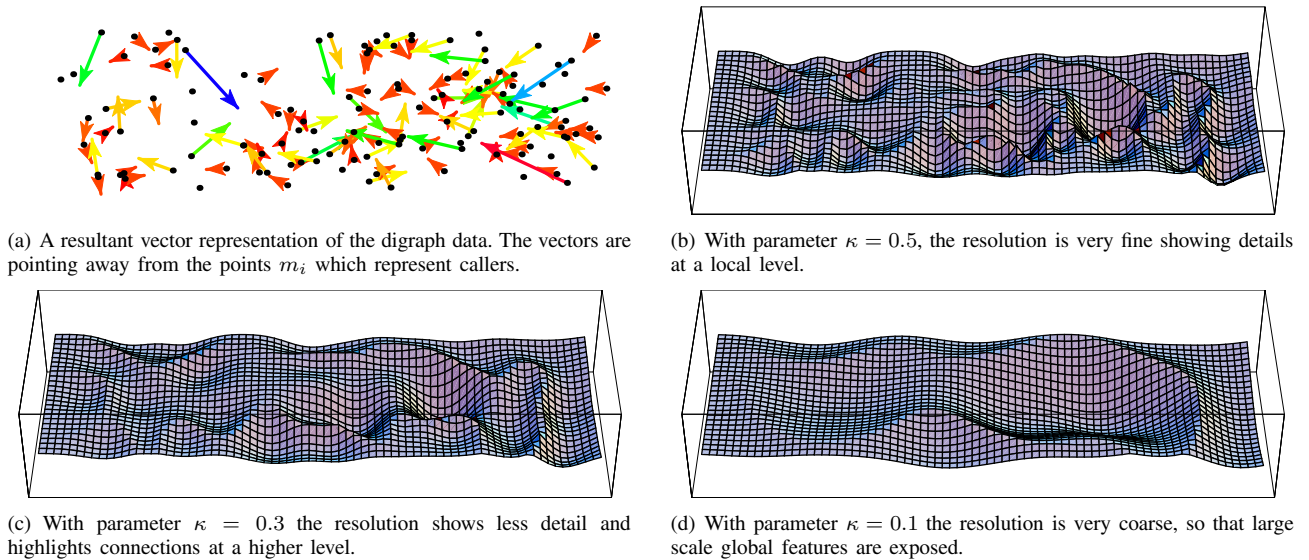
Figure 5.   Visual representations of a hypothetical community of 124 callers who collectively have made 223 calls. Figures 5(b), 5(c) and 5(d) are the negative absolute values of the torque surface defined in Definition II-B for different values of κ [14], [15].

C. Examining the effect of the parameter κ

The κ parameter controls the extent of the distortion effect. A larger κ will result in a smaller region of distortion around each m_r point which has a corresponding ν_r with non-zero magnitude. For real data where n is large, the parameter κ can be "tuned" to reveal areas of interaction at different scales; large κ would reveal details at a local neighborhood level; while increasingly smaller κ would show details at the town, county, state or nation levels.

An example of a hypothetical community of 124 callers who collectively have made 223 calls can be observed in Figure 5. Figure 5(a) displays the data as resultant vectors. The surfaces in Figures 5(b), 5(c) and 5(d) have been displayed as negative absolute values (the mountains have been turned into valleys) for different values of κ (different resolutions).

This was done to highlight what is called a clustering effect. The ridges that encircle a depression are considered as geographic region of network nodes, (a cluster), that generally share a property at a particular scale dependent on κ. In general, this shows that the resultant vectors within the region are distributed in such a way as to be oriented towards the center of the region. It can be seen in Figure 5 that as κ decreases larger scale clusters appear [14], [15].

482

REFERENCES

[1] James Abello, Jeffrey Korn, Matthias Kreuseler, Navigating Giga-Graphs, *AVI 2002 Advanced Visual Interfaces*, Trento, Italy, pp. 22–24, 2002.

[2] James Abello, Panos M. Pardalos and Mauricio G.C. Resende, On Maximum Clique Problems in Very Large Graphs, in External Memory Algorithms, J. Abello and J. Vitter (Eds.), *AMS-DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, Vol. 50, 1999.

[3] James Abello, Panos M. Pardalos and Mauricio G.C. Resende, (Eds.), *Handbook of Massive Data Sets*, Book Series: Massive Computing : Volume 4, Kluwer Academic Publishers, Dordrecht, 2002.

[4] James Abello, Shankar Krishnan, Navigating Graph Surfaces, *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, Panos Pardalos (Ed.), Kluwer Academic Publisher, Dordrecht, Boston, London, pp. 1–16, 1999.

[5] James Abello and Jeffrey Korn, Visualizing Massive Multi-digraphs, *IEEE Information Visualization 2000*, Salt Lake City, UT, pp. 39–48, 2000.

[6] James Abello and Jeffrey Korn, MGV, A System for Visualizing Massive Multi-digraphs, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 8, No. 1, pp. 21–38, 2002.

[7] James Abello, Irene Finocchi and Jeffrey Korn, Graph Sketches, *IEEE Symposium on Information Visualization*, IEEE Computer Society, 2001.

[8] Tamara Munzner, Interactive Visualization of Large Graphs and Networks, http://graphics.stanford.edu/papers/munzner_thesis, Thesis, accessed May 2004, 2000.

[9] David Harel and Yehuda Keoren, A Fast Multi-Scale Method for Drawing Large Graphs, *TR MCS99-21*, The Weizmann Institute of Science, Rehovot, Israel, 1999.

[10] Aaron J. Quigley, Force Directed Algorithms by Decomposed Estimation, http://www.cs.usyd.edu.au/~aquigley/3dfade/, accessed June 2004.

[11] Pawel Gajer, Michael T. Goodrich and Stephen G. Kobourov, A Multidimensional Approach to Force Directed Layouts of Large Graphs, Proceedings of Graph Drawing, *Lecture Notes of Computer Science*, Springer Verlag, 2000.

[12] Robert Spence, Information Visualization, *Addison-Wesley*, 2001.

[13] Method of Data Processing and Representation http://www.wipo.int/pctdb/en/wo.jsp?wo=2007053910

[14] Victor Korotkikh and Noel Patson, A New Approach to Analyze Huge Internet and Telecommunications Data, *International Conference on Advances in the Internet, Processing, Systems and Interdisciplinary Research*, CAITA-2004 Purdue, Conference Proceedings, Purdue, Indiana, U.S.A., 2004.

[15] Noel Patson, A Decomposition Framework for Visualising Telecommunications and Cyberspace Systems, PhD Thesis, Central Queensland University, 2005.