Eurographics Conference on Visualization (EuroVis) (2014)
N. Elmqvist, M. Hlawitschka, and J. Kennedy (Editors) *Poster*

Influent: Scalable Transactional Flow Analysis with Entity-Relationship Graphs

David Jonker, Scott Langevin, David Gauldie and William Wright[1]

[1]Oculus Info Inc., Toronto, Ontario, Canada

Abstract

In this paper, we present in-progress work on "Influent", a graph analysis tool that enables an intelligence analyst to visually and interactively "follow the money" or other transaction flow. Summary visualizations of transactional patterns and entity characteristics, a left-to-right semantic flow layout, interactive link expansion and hierarchical entity clustering enable Influent to operate effectively at scale with millions of entities and hundreds of millions of transactions, with larger data sets in progress.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Computer Graphics]: Graphical User Interfaces (GUI)—

1. Introduction

Challenges with using graphs (node-link diagrams) to visualize extremely large entity-relationship datasets include visibility, usability and high degree nodes. "Hairballs" resulting from trying to portray even just a fraction of such datasets are difficult and time consuming to explore and understand [SA06]. Analytical questions such as determining node connectivity and behaviors over time are challenging.

Our motivation is to develop a tool to assist forensic analysts in understanding transaction flow amongst entities and to identify trends, patterns, suspicious entities or transactions and anomalous behavior. Our tool, Influent, is designed for interactive graph analysis tasks to explore entity relationships, characteristics and transactional behaviors that perceptually and computationally scale to very large datasets. To prototype our approach, we tested Influent using publicly available data for the financial forensics domain. The dataset used was obtained via the Kiva [Kiv14] non-profit microfinance organization API and comprises 2 million accounts and 150 million transactions.

2. Large Scale Transaction Graphs and Analysis

Related research has attempted to address sense-making and interactive graph exploration of large graphs [LPP*06] [vHP09] [HB05] [SA06] [BDL*10] [HFM07]. Apolo [CKHF11] has similar goals of constructing bottom up views of large networks through interactive expansion of network links and incorporating machine learning. However, previous research has limited support for interactive analysis of temporal graphs of transactional links or aggregate visualization to summarize clusters of nodes. Our work aims to address these deficiencies.

3. Design for Perceptual Scalability

Influent addresses perceptual scalability through the use of on-demand expansion of graph links, hierarchical clustering of graph entity nodes, aggregation of transactional links and aggregation markers. Due to the size of the entity graphs, and their transactional nature, displaying all node and links was not feasible to support meaningful sense making. Our approach is flow-oriented navigation of the graph, by allowing for interactive link expansion of the graph view to display inbound or outbound links centered on a set of entity nodes. This allows for user exploration of relevant sub-networks of entities. Semantic layout of nodes and links represents flow from left to right (Figure 1).

To overcome loss of information common with aggregation, Influent uses aggregation markers that provide a visual summarization of the aggregation (Figure 2). Visual aggregates convey additional information about the underlying contents, such as histograms that summarize temporal behaviors and distributions of attributes.

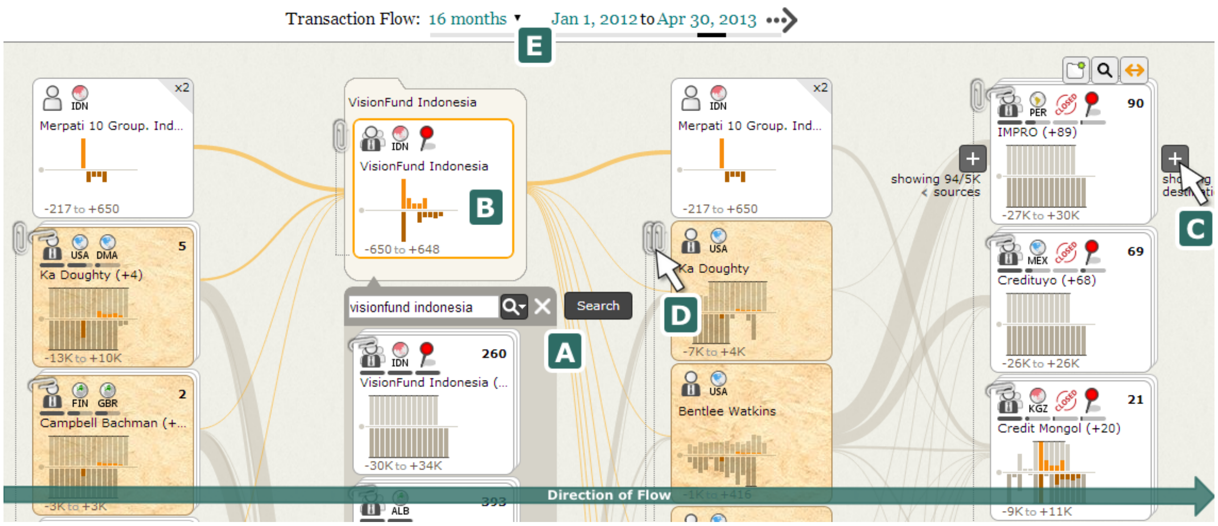

Figure 1: *Example of Influent flow analysis with the Kiva dataset. A user begins by searching for entities of interest (A). Entity cards summarize entity properties and histograms charts summarize transactions for the specific time filter. A file metaphor is used for users to drag search results into a workspace for exploring their transaction graph. An entity can be designated as the focus entity where associated transactions and links are highlighted (B). Click [+] buttons to expand incoming/outgoing entity links (C). Large numbers of entities are hierarchically clustered and can be incrementally unstacked by clicking paperclip to drill down (D). Transaction data can be filtered using the transaction time filter (E).*

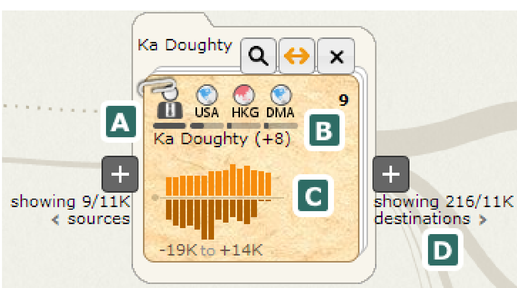

Figure 2: *Example Influent aggregate markers summarizing entities. Clustered entities are distinguished with a paper clip and stack (A). Property distributions (B) summarize entity properties such as location and entity type. Time series (C) summarize transaction activity and highlight flow to and from entities of interest. Degree markers (D) indicate how many inflowing or outflowing entity connections exist and displayed. Link widths summarize "flow" across links.*

3.1. Hierarchical Ensemble Clustering of Entities

Visual scalability issues arise when expanding links from high degree nodes. High degree nodes have a large number of inbound or outbound links to other entities. To overcome visual clutter, we applied hierarchical clustering of entities (shown with a stacked cards metaphor) and aggregated links (shown with a Sankey-like visual expression). This approach adaptively splits clusters of entities into sub-clusters to reduce visual scale and allow drill-down exploration of the clusters. Our scheme is an ensemble clusterer, which iteratively clusters entities into sub-clusters using a configurable set of entity attributes. Each conceptual level in the cluster hierarchy represents a clustering of entities by a different attribute to aid in meaningful navigation.

4. Conclusions and Future Work

Future work is planned for interactive exploration such as filtering links of interest, and link search operations. The left-to-right placement of entities can result in duplicate entities being displayed when flow are coming back to their origin. Influent marks these entities with a "duplicate" label (see Merpati entity in Figure 1). Alternate strategies will be explored. Lastly, formal user experimentation should be conducted to study the effectiveness of Influent methods. A live demo of Influent configured for the Kiva dataset is hosted at `http://www.influent.org`.

5. Acknowledgements

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Contract Number FA8750-12-C-0317. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision.

References

[BDL*10] BATAGELJ V., DIDIMO W., LIOTTA G., PALLADINO P., PATRIGNANI M.: Visual analysis of large graphs using (x,y)-clustering and hybrid visualizations. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE* (March 2010), pp. 209–216. 1

[CKHF11] CHAU D. H., KITTUR A., HONG J. I., FALOUTSOS C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 167–176. 1

[HB05] HEER J., BOYD D.: Vizster: visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on* (Oct 2005), pp. 32–39. 1

[HFM07] HENRY N., FEKETE J., MCGUFFIN M.: Nodetrix: a hybrid visualization of social networks. *Visualization and Computer Graphics, IEEE Transactions on 13*, 6 (Nov 2007), 1302–1309. 1

[Kiv14] Kiva, 2014. URL: `http://kiva.org`. 1

[LPP*06] LEE B., PARR C., PLAISANT C., BEDERSON B., VEKSLER V., GRAY W., KOTFILA C.: Treeplus: Interactive exploration of networks with enhanced tree layouts. *Visualization and Computer Graphics, IEEE Transactions on 12*, 6 (Nov 2006), 1414–1426. 1

[SA06] SHNEIDERMAN B., ARIS A.: Network visualization by semantic substrates. *Visualization and Computer Graphics, IEEE Transactions on 12*, 5 (Sept 2006), 733–740. 1

[vHP09] VAN HAM F., PERER A.: Search, show context, expand on demand: Supporting large graph exploration with degree-of-interest. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (Nov 2009), 953–960. 1