

# Hamlet: a Metaphor for Modeling and Analyzing Network Conversational Adjacency Graphs

Joseph P. Macker\*

\*Information Technology Division, Naval Research Laboratory, Washington DC

Email: joseph.macker@nrl.navy.mil

**Abstract**—This paper presents a working model of conversational adjacency graphs (CAGs), both temporal and non-temporal, useful in analyzing interactive communication patterns amongst distributed network actors. The paper utilizes Shakespeare’s Hamlet as a baseline reference model to outline the construction of representative weighted graph models of inter-actor conversations, both temporal and non-temporal. A sampling of related complex network analytic metrics both temporal and non-temporal variants is also presented on the respective graph models. We discuss a related experiment carried out in which network exchanges amongst distributed actor nodes are orchestrated within an emulated virtual wireless network scenario. A *blind* construction of the representative CAG is presented from received traffic logs without a-priori knowledge of the play ordering. In this case, source identification and reception ordering are the sole pieces of information used to estimate conversational adjacency relationships. We demonstrate that the example “blind” CAG construction from the empirically-measured model appears to reasonably match the baseline conversation graph model developed with a-priori information directly from play script. Issues and related ongoing work is discussed regarding the related analytics and potential applications of such models.

## I. INTRODUCTION

Random network traffic patterns are used throughout a significant portion of past ad hoc wireless network research as a means to establish comparative statistical measures between system designs. While approaches remain useful for generic baseline studies of systems, there is a growing interest and need to focus on more realistic workflow-based network traffic models. How will the network actually be used and what is its effectiveness in regard to its unfolding support of a mission(s)? There is also dual interest in better understanding and modeling communication exchanges and emergent causalities that may exist amongst complex network services and actors within actual communication network deployments. Due to complexity and the ongoing abstraction of planned networked systems, we have a lack of effective means to model and/or measure temporal communication adjacency relationships and how they evolve in time. This paper focuses in three main related areas: First, it presents a canonical model for time-ordered communication messages as an estimation of conversational relationships. Second, it addresses and compares the formulation of analytic conversational graph(s) from both an a-priori model (e.g., play script or workflow) and a experimentally measured model (e.g., from captured network data). Third, it presents structural analytic examples using applied

complex network theory for related aggregate and temporal graph models.

In this work, we discuss the capability of structurally representing such systems using time-windowed graphs to enable analysis and characterization of roles and interactions amongst key network players. Albeit it static and pre-existing, William Shakespeare’s Hamlet [1] is a well-established, time-ordered story of multiple characters and their various interactions. In this sense, Hamlet serves as an example to formulate weighted graph models of directed conversational adjacencies and later in the paper we present results from an actual orchestrated network experiment of conversations amongst distributed network actors. From captured network emulation data, we construct graph models without a-priori knowledge of the play by using source identifiers and reception ordering information. We compare these results against a-priori aggregate and temporal graph models of dialogue interactions. We also discuss potential applications of complex network analytic techniques to perform statistical inference regarding various actor relationships and roles. Our research goal is the application of similar techniques and analysis methods to less pre-determined network collaboration and workflow systems.

## II. PARSING HAMLET INTO CONVERSATIONAL ADJACENCIES

To start, a baseline a-priori graph model of Hamlet conversations is extracted from the freely available text version of the play at Project Gutenberg [2]. We use developed software tools to pick out temporally ordered events such as: Acts, Scenes, dialogue, and other actions. Figure 1 shows an overview of the stored data structure including the order and assignment of speaking roles. Key events throughout the play are well tagged supporting the ability to distinguish Acts, Scenes, Actions, and dialogue stanzas. In some infrequent cases, multiple actors are directed to speak as a *Chorus*, a simultaneous group event, and in these cases we assign dialogue broadcast events to a subset of multiple source actors based upon the scene context of the play.

## III. WINDOWED CONVERSATIONAL ADJACENCY GRAPH MODEL

From an ordered set of dialogue events, we develop a reference graph-based representation we refer to as a conversational adjacency graph (CAG). The formulation is consistent with past applied contact graph models discussed within [3],

although a difference is that when applied to actual logged network data we can optionally infer a pair-wise temporal adjacency from the order of dialogue events and not from any specific pair-wise receiver contact information. If such specific information is available, as in email traffic or other application or network data, one can improve the fidelity of the causal representation of the model by using it, but not requiring specific destination data enables the use of the model in situations such as multicast group collaboration scenarios where pair-wise causal relationships may not be obvious. One relevant application of such models is to better understand group-based workflows or unfolding interactions within event-driven network communication scenarios, especially as we begin deploying more distributed network services and collaborative applications. Towards the end of this paper, we present

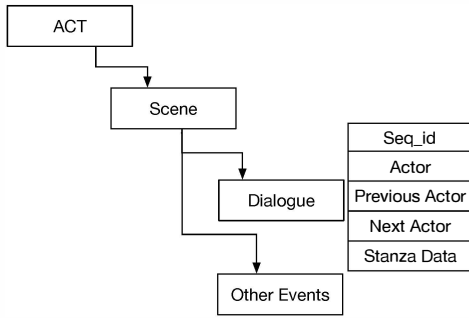


Fig. 1: Play Related Structure

results from a network testbed experiment in more detail. For now, we focus on an a-priori model and more formally define the time-windowed CAG construction and related analytic applications. To develop an a-priori CAG model, we use the dialogue sequence identifiers and the information in previous and next actor entries from Figure 1. This data also later gives us sufficient information to orchestrate actor actions within our distributed network experiment.

We define a  $CAG_{t_n, t_k}$  as a time-windowed digraph in the form  $G(V, E, w_v, ew_{i,j})$ , with  $V$  vertices, and  $E$  represented directed edges such that:

- 1) Any  $v \in V$  represents an actor as a weighted vertex within the conversation group.
- 2)  $w_v$  represents actor's vertex weight in terms of conversational activity during  $t$ , such that  $t_n \leq t \leq t_k$ .
- 3)  $E$  represents the edges of the CAG as a directed conversational adjacency between two actors. A directional edge  $e_{i,j} \in E$  exists if  $v_i$  speaks and  $v_j$  follows during  $t$ , such that  $t_n \leq t \leq t_k$ .
- 4)  $ew_{i,j}$  represents edge weight in terms of directed pair-wise conversational activity during  $t$ , such that  $t_n \leq t \leq t_k$ .

Weights within CAGs, both  $w_v$  and  $ew_{i,j}$ , represent either

some event activity count (e.g., number of conversation events) or an absolute aggregate amount of directed conversational data (e.g., conversational bytes transmitted). Within our canonical model, the Act and Scene changes also make up exceptions in estimating  $e_{i,j}$  relationships if a windowed  $t$  duration overlaps Scenes and Acts. At the beginning of an Act or Scene change the first actor is likely speaking autonomously and starting off a conversation and at the end of a Scene the final speaker is likely ending a conversation thread so a directed adjacency is excluded in these cases in the a-priori model. Within Scenes, conversational adjacencies are only estimators of potential conversational causality but in many circumstances (e.g., chat rooms) its reasonable to assume some degree of conversational causal relationships exist given a large enough set of data. As mentioned, if additional intended receiver information is available this adds fidelity to construction of the model.

From conversational data, either a-priori known or *blind* estimated from network ordering, we build up representative graph structures and weights to represent the conversational events. If we are using adjacency events as the metric, each event is weighted as 1 irregardless of length. An example of  $w_v$  variations is shown below for Francisco who speaks only in Scene I, Act I. The first value in the data tuples shown is a global sequence identifier within the play and the second value represents the length in bytes of the referenced actor's stanza.

```
print my_diag_ids['Francisco']
```

```
[[1, 46], [3, 12], [5, 42],
 [7, 73], [9, 24], [11, 49],
 [14, 23], [16, 46]]
```

In the aggregate CAG case,  $t$  = the entire play, if event count is chosen as the vertex weight method, we ignore the length values and  $w_v = 8$ , but if the amount of transmitted data is chosen, then  $w_v = 315$ . This same approach of event count vs. information quantity follows for potential choices of directed edge weights,  $ew_{i,j}$ . In all further model examples and experiments within this paper, we apply event count exclusively as the metric for both node and edge CAG weights although the model directly supports modification for additional weighted measures such as total bytes in related stream exchanges amongst actors.

#### IV. INITIAL CAG CONSTRUCTION AND ANALYSIS

Figure 2 represents a compressed ordered event view of Hamlet without stanza length information included. In developing our reference aggregate CAG example,  $CAG_{t_n, t_k}$ , we use this entire timeline to construct a directed, weighted graph representation. For this CAG, vertex weights represent the number of conversation events and edge weights represent the number of directed pair-wise conversational adjacencies. Figure 3 presents a visualization of resulting the graph structure with relative actor event rankings indicated by node size and edge widths representing pair-wise conversational adjacency events. To improve the visualization readability directional edge weights have been removed from the graph.

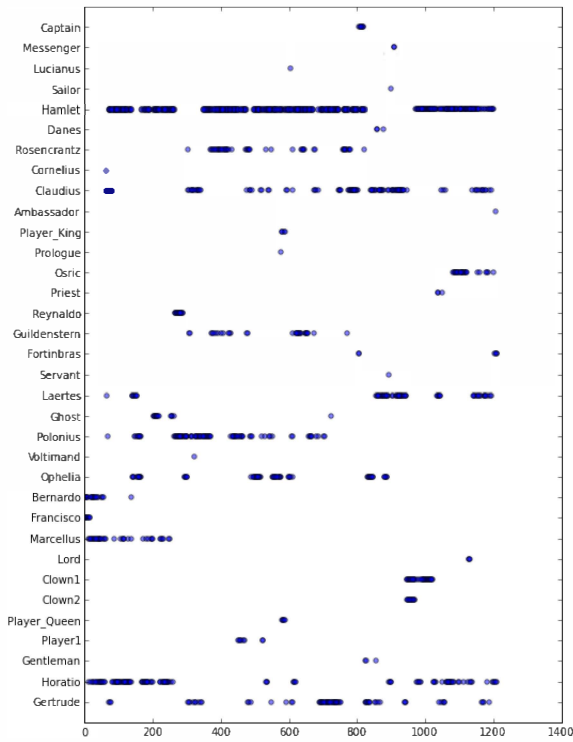


Fig. 2: Hamlet Dialogue Events by Actor

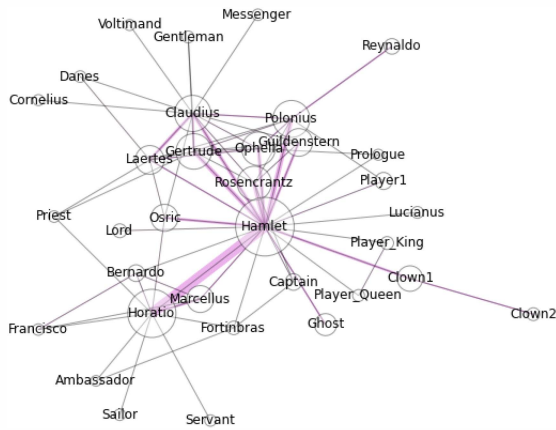


Fig. 3: Hamlet CAG Overview

From the basic aggregate CAG structure, we now produce Figure 4 representing a histogram of the top 9 dialogue events by actors. This is basically a plot of the  $w_v$  vertex weight values of the CAG. Further, based upon directed edge information in the CAG we show the results from sorting and plotting the top 9 adjacencies throughout the play in Figure 5. It is perhaps not surprising in Figure 5 that the two top adjacencies for the aggregate CAG are between Hamlet and his faithful friend Horatio. Finding out such information in the blind estimation case later demonstrates some of the usefulness of this approach.

Of course, we could have produced the above results (e.g., dialogue histograms) without bothering to build a CAG itself, but by formulating a CAG model we now have a significant body of analytical work from complex network theory and

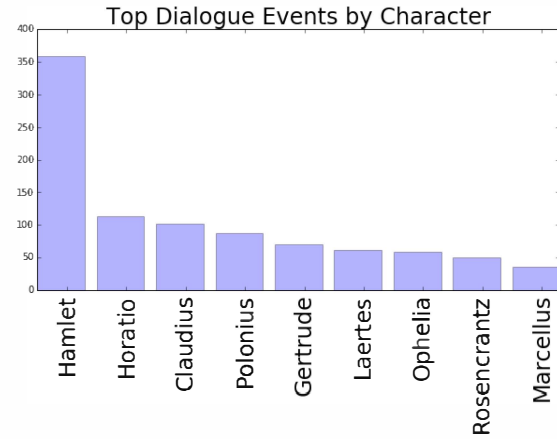


Fig. 4: Hamlet Dialogue Event Histogram (top9)

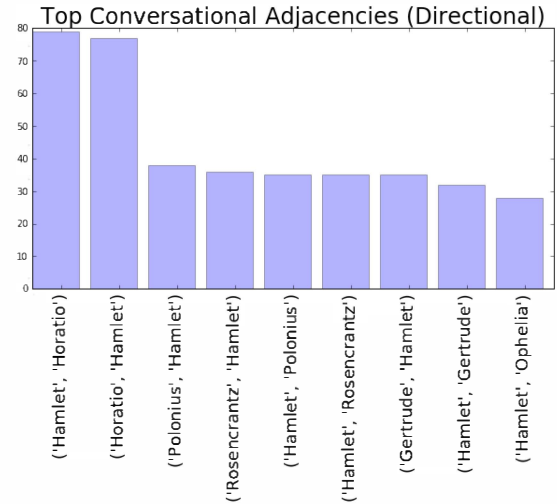


Fig. 5: Hamlet Top Conversation Adjacencies

graph-based analysis of structure at our disposal. Related analytics can provide more advanced insight into complex structural relationships (e.g., clustering) and temporal models may further reveal insights regarding different phases of collaborative mission exchanges or workflows. While applied complex network theory is an actively evolving field, present applicable analytical metrics include community clustering, global structural invariants, and statistical centrality metrics for both nodes and edges. One popular metric approach is the use of centrality measures representing statistical rankings of the importance or influence of vertices (i.e., nodes) or edges based upon a particular structural or interaction model [4], [5]. Several centrality measures have mathematical foundations in statistical mechanics and can therefore be potentially useful in predicting forms of node interaction and information flows with a structure. Examples include the ability to estimate node or edge betweenness of conversational threads (e.g., varieties of betweenness measures). We can also apply results from complex network theory to examine global structural invariants like whether CAGs exhibit assortativity properties in which actors with high conversation transmission or reception tend to connect to other like actors. Due to space limitations, we demonstrate only a few potential analytic examples using the

a-priori aggregate Hamlet CAG model.

**Global Invariant Metric Example: Directed Assortativity Measure:** In a directed graph, in-assortativity  $r_{in}$  and out-assortativity  $r_{out}$  measure the tendencies of nodes to connect with other nodes that have similar in and out degrees as themselves and are defined in [6] as Eq. 1.

$$r_{in} = \frac{1}{\sigma_q^{in} \sigma_q^{in}} [(\sum_j k e_{j,k}^{in}) - \mu_q^{in} \mu_q^{in}] \quad (1)$$

$e_{j,k}^{in}$  is the joint probability distribution of links going into target nodes with  $k$  out-degrees, and out of source nodes of  $j$  out-degree.  $q_k^{in}$  is the probability distribution of links going into target nodes with  $k$  in-degrees.  $\hat{q}_j^{in}$  is the probability distribution of links going out of source nodes with  $j$  in-degrees.  $\sigma_q^{in}$  and  $\mu_q^{in}$  are respectively the mean and standard deviation of  $q_k^{in}$ . Similarly  $\sigma_{\hat{q}}^{in}$  and  $\mu_{\hat{q}}^{in}$  for  $\hat{q}_k^{in}$ . See [6] for similar definition of  $r_{out}$ .

We obtain the following results for both weighted and unweighted variants of the directed baseline CAG.

- $r_{in}$  (weighted) = -0.28
- $r_{out}$  (weighted) = -0.28
- $r_{in}$  (unweighted) = -0.37
- $r_{out}$  (unweighted) = -0.37

The above negative results indicate a partially disassortative nature to the overall set of conversational relationships in Hamlet. Upon a rough review of the play, there are indeed plenty of significant conversations Hamlet and other major event weighted actors carry on directly with more minor characters (e.g., Osric, Players, etc) so this metric seems to give insight into that behavior.

**Node and Edge Centrality Example:** We next briefly present a more sophisticated complex network analytic view of the aggregate CAG relationships using simultaneously a node-centric current flow betweenness measure and a conversational edge-centric betweenness measure. To represent betweenness centralities properly in a weighted graph computation we need an additive cost metric available for measuring shortest cost paths and we use a reciprocal of edge weight values for cost/distance. We review the definition of these measures in the non-temporal graph case. Edge Betweenness Centrality is defined in [7] as Eq. 2.

$$E_{bet}(e) = \sum_{s,t \in V} \frac{\sigma_{st}|e|}{\sigma_{st}} \quad (2)$$

$\sigma_{st}$  is the total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_{st}|e|$  is the number of those shortest paths that pass through edge  $e$ . This is a straightforward variation of node betweenness within a graph but applied to edges.

The basic current flow betweenness centrality,  $C_i^{RWB}$ , also known as random walk betweenness, is defined in Equation 3

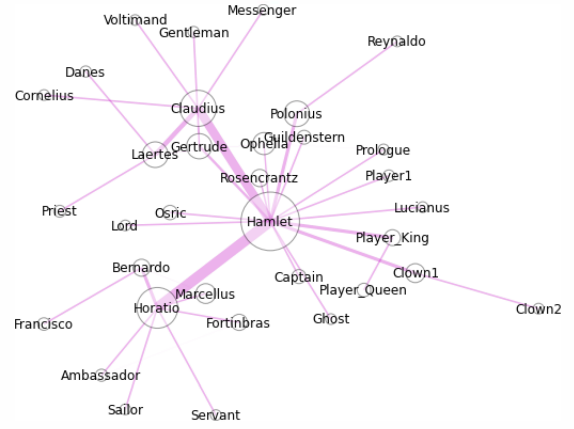


Fig. 6: Hamlet Flow Analytics

and represents a view of the graph modeled as an electric circuit model, i.e. Kirchhoff's laws [8]. An interesting feature of this model over more frequently applied shortest path betweenness centrality is its accounting for the contribution multiple information path flows.

$$C_i^{RWB} = \sum_{j \neq i \neq k} r_{jk} \quad (3)$$

where the  $r_{jk}$  element of matrix  $R$  contains the probability of a random walk starting at node  $j$ , that absorbs node  $k$ , passing through node  $i$ .

An analysis of the Hamlet CAG representing node size as  $C_i^{RWB}$  and edge size as  $E_{bet}(e)$  is shown in Figure 6. An observation from this analysis is that the top two  $E_{bet}(e)$  results are *Hamlet – Horatio* and *Hamlet – Claudius* edges. This provides some additional analytical insight in that *Hamlet – Horatio* was already in the previous top activity edges and remains as a high betweenness edge, i.e., bridging other conversations or serving as an important information dissemination edge. *Hamlet – Claudius* had a lower metric than several others on the purely "activity" weighted ranking but now ranks as the second strongest result using an edge betweenness metric. Again, this serves as an indicator that conversations between *Hamlet – Claudius* are frequently on the shortest path between other conversational adjacencies within the graph. By studying the play and its unfolding plots, we can see how the *Hamlet – Claudius* is a strong structural conversational bridge between other actor conversations as these actors speak together many times but also interact with many separate characters, frequently in the absence of each other. Such analytical modeling and insight into complex structural relationships is a short example of key applied research goals for ongoing work we are considering as such behavior could represent collaborative but separately clustered communication leaders or the behavior of distributed proxies.

## V. TEMPORAL CAG MODEL EXTENSIONS AND APPLICATIONS

The basic aggregate CAG model and example has value for numerous applications but there is strong motivation to develop time-ordered causality representations better supporting evolving interactions or time-oriented information spreading statistics. By developing graph representations preserving the temporal ordering of conversation adjacencies we can go deeper to apply innovations in complex network metrics for temporal graphs [3], [9], [10]. This type of temporal modeling requires either a directed multigraph representation (with time information recorded on edges), or equivalently, a time-ordered series of digraphs. Building upon the time-windowed CAG definition, we now discuss applied work using temporal models of CAGs (directed or undirected). We have developed software tools to construct these graph sequences from data sets and to perform a variety of temporal metric calculations from recent literature. A basic benefit of a time-ordered graph model is that communications with causal or time-ordered structures are better represented (e.g., temporal disconnections, arrival/departures of actors, temporally-ordered workflow exchanges). A straightforward example in our present canonical model is the actor, Francisco, who appears and transacts only in Scene 1, Act 1. Therefore, this actor has no potential temporal information propagation back to him from actors appearing later in the play (e.g., Rosencranz). However, in the previous aggregate graph formulation, such a graph pathway, albeit minor, exists between Rosencranz and Francisco through Hamlet since communication time-ordering was not preserved in the model.

### VI. SCENE-BASED, TIME-VARYING CAGS

To provide an example of temporal graph analytics for Hamlet, we construct a time-ordered series of CAGs representing windowed CAGs on Scene boundaries. We could carry this filtered of CAG relationships down to smaller time windows as needed, and at a lower limit we have an ordered series of windowed graphs representing a single conversational event in each graph. A Scene within a play serves as a natural context for a series of conversation exchanges and real world mission modeling of chat rooms or other temporal collaborations may have similar contextual time or group filters related to mission phases or subtask collaborations. Each specific application will likely require appropriate consideration of the time-windowing implications and relevant estimations (e.g., time clustering) but the generic model supports flexibility in this regard.

### VII. SCENE-BASED HAMLET CAGs: TEMPORAL METRIC EXAMPLE

A temporal geodesic betweenness metric was developed in [11] and we show related results from our time-ordered analysis of all Scene CAGs in Figure 7. This shows the top 5 nodes across temporal graphs that are "in between" other nodes in terms of temporal conversational distance or cost. As a quick example of temporal filtering effects when we reran this data for Scenes 1-3, only Polonius has a positive value

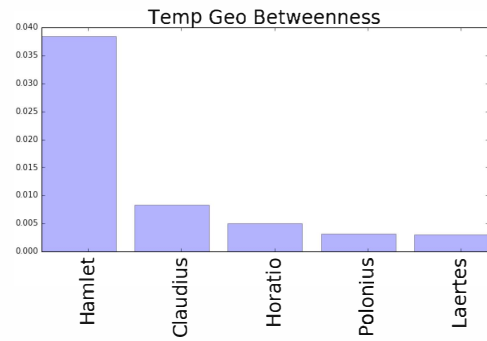


Fig. 7: Top 5 Temporal Geodesic Betweenness Nodes

who is the only common actor residing between these Scenes. Due to space limitations, we leave further temporal metric analysis and discussion to further work.

## VIII. BLIND CAG CONSTRUCTION FOR EXPERIMENTAL NETWORK DATA

To test out the basic construction and analysis of a CAG from experimental data, we orchestrated a network emulation experiment using distributed network-based Hamlet actors. The basic experiment model was constructed using 35 virtual network containers within the COMmon Research Emulator (CORE) [12]. A single actor was assigned to a network node and conversational activity was orchestrated using the *basicmgenactor.py* remote interface for MultiGenerator (MGEN) [13] network traffic test tool over the emulation control interface. This experiment focuses on generating text messaging from actors (e.g., chat group) although we have also orchestrated the network streaming of relevant audio sound bytes. To construct a CAG estimate, we used received network traffic logs at a node designated as a *mock* chat server within the emulated network. Since this CAG estimate model is independent of receive identifiers, a companion multicast network test case resulted in similar findings. Our unicast case is similar to a centralized chat server type deployment where a log may be available of incoming session traffic. In the multicast case, we might have an application model where all sources or a pub/sub message bus uses a well-known multicast destination and this could be used to monitor sequenced collaborative interactions. In the simplest case, we use only the order of arriving stanzas and a source identifier to construct the conversation adjacency estimates.

Figure 8 presents our estimated aggregate weighted CAG from the unicast-based network experiment and Figure 9 presents a list of the top 10 sensed aggregate conversational adjacencies from this actual experimental data. There is strong agreement between this result and the a-priori model (e.g., the top eight adjacency rankings agree). There some minor differences in rankings for subsequent edges (e.g., 9 -10) but this is likely explained by the absence of Scene change estimate in the empirical example. Strong bi-directional relationships between *Hamlet – Horatio*, *Hamlet – Polonius*, *Hamlet – Rosencranz*, *Hamlet – Gertrude* are detected as in the a-priori model purely from this *blind* post analysis.



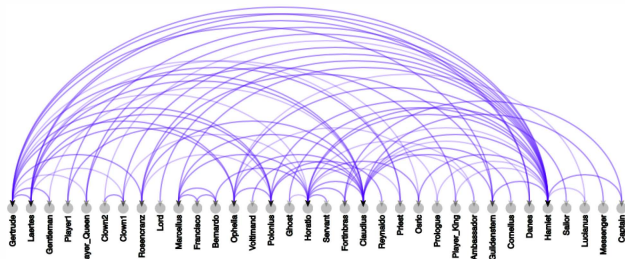


Fig. 8: Arc Diagram of Blind Hamlet CAG

This is relatively trivial example but demonstrates the concept in action within an actual network testbed.

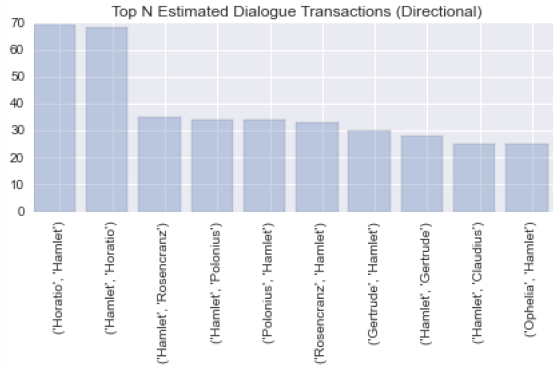


Fig. 9: Top 10 Directional Adjacencies from Blind Construction

## IX. ISSUES AND ONGOING WORK

We are actively working on applying CAG models for modeling and analyzing collaborative tactical network services and distributed workflows. A research goal is improving the measurement and analysis of evolving network collaborative applications and transactional exchange structures. These same capabilities are presently assisting in the modeling and orchestration of distributed causal systems within ongoing network experiments. We are carrying out ongoing work into temporal graph models and at present we have implemented and experimented with temporal variants of centrality including: communicability, broadcast, receive, geodesic betweenness, transitive betweenness. We should point out that while important, time-varying graph models add complexity and are not an analytic panacea. At present, there is no *one size fits all* approach. As an example, some temporal metric formulations ignore or underemphasize the importance of communication pathways within each time window and focus only on temporal connections and pathways between graph sequences. Many tactical communication network models evolving in time are easily able to exchange local data within subgraphs on shorter timescales so we feel it is important to analyze connected graph pathways within the time windowed CAGs (e.g., short term exchanges) as well as to consider segments necessary to propagate information in time between sequenced CAGs (e.g., store-forward propagation model). In this regard, we are looking into models preserving intra-sequence pathway information similar to the model outlined by (Tang, et al)

in [14]. Recent theoretical constructions, such as temporal communicability betweenness [10] also provide some ability to emphasize different temporal timescales and we are actively examining such approaches.

## X. CONCLUSIONS

We presented a model of a time-windowed CAG and demonstrated construction and potential applications using dialogue sequence and actor information from Shakespeare's Hamlet. We provided samples of complex network analytics on the CAG model as a canonical example of examining conversational relationships as they might occur between applications or collaborative services within a distributed network. Distributed group chat, collaborative distributed computing, and mission-based network workflows are some examples of potential real world applications of DoD interest. We presented uses of the model first as an aggregate, time-independent analytical structure and we extended this to a set of scene-based, temporal graph models of an evolving nature. We presented a short discussion of initial temporal graph-based metric results and discussed further planned investigations. Lastly, we presented an empirical example of constructing a CAG model from an actual network emulation without a-priori knowledge of the actor relationships or conversations. Results demonstrated that we could construct a CAG model largely agreeing with an a-priori CAG model and could gain insight into the structural role of nodes and edges.

## REFERENCES

- [1] William Shakespeare. *The Tragedy of Hamlet*. University Press, 1904.
- [2] Marie Lebert. Project Gutenberg (1971-2008), 2008.
- [3] Vincenzo Nicosia, John Tang, Cecilia Mascolo, Mirco Musolesi, Giovanni Russo, and Vito Latora. Graph metrics for temporal networks. In *Temporal Networks*, pages 15–40. Springer, 2013.
- [4] Mark EJ Newman. The Structure and Function of Complex Networks. *SIAM review*, 45(2):167–256, 2003.
- [5] Ernesto Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, 2011.
- [6] Piraveenan et al. *Assortative mixing in directed networks*. University of Sydney, School of IT, 2009.
- [7] Ulrik Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.
- [8] Ulrik Brandes and Daniel Fleischer. *Centrality measures based on current flow*. Springer, 2005.
- [9] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.
- [10] Ernesto Estrada. Communicability in temporal networks. *Physical Review E*, 88(4):042811, 2013.
- [11] H Habiba, C Tantipathananandh, and T Berger-Wolf. Betweenness centrality measure in dynamic networks. *Department of Computer Science, University of Illinois at Chicago, Chicago*, 2007.
- [12] Ahrenholz et al. Core: A real-time network emulator. In *Military Communications Conference*, pages 1–7. IEEE, 2008.
- [13] Brian Adamson. MGEN User's and Reference Guide Version 5.0, <http://downloads.pf.itd.nrl.navy.mil/docs/mgen/mgen.html>, March 2016.
- [14] John Tang, Mirco Musolesi, Cecilia Mascolo, and Vito Latora. Characterising temporal distance and reachability in mobile and online social networks. *ACM SIGCOMM Computer Communication Review*, 40(1):118–124, 2010.