Measuring 16 Years' Evolution of a Collaborative Water Planning Network

Abstract

To enable collaborative, networked governance approaches to be optimally effective, it's vital to understand how to create and sustain high-performing networks over their full lifecycle. This paper observes the evolution of the governance network involved in an externally mandated, highly collaborative process for licensing a hydropower dam in Washington state. We introduce a novel approach using text mining techniques and temporal statistical network analysis to observe meeting participation networks from meeting minutes for a 16-year period. We find that the network was remarkably stable over the 16-year period, with network conveners and other individuals who played leadership roles early in the process likely to remain in those roles over time. More active participants (measured as individuals who were speaking during meetings) were more likely to sustain engagement. Additionally, partway through the 16-year period, the external mandate to collaborate dissolved, leading to a more tumultuous network with higher turnover. These findings suggest that high-performing mandated networks evolve quite differently than grassroots networks, with much less turnover and a strong reliance on a constrained group of conveners to sustain interaction.

Keywords: network governance, network evolution, TERGM, natural language processing, water planning and management, FERC hydropower relicensing

Introduction

Collaborative, networked governance approaches have become a core feature of many planning and administrative processes, growing in popularity as they help organizations share resources more efficiently to address complex governance challenges. However, understanding how to create and sustain these networks over time is an open question. Building effective networks requires knowledge about how to incentivize participation (Sabatier et al. 2005), supply the resources and leadership necessary to support engagement (Emerson and Nabatchi 2015a), and adopt context appropriate governance structures (Provan and Kenis 2008). Because networks are dynamic, constantly shifting in character to reflect varying management goals and participants, what the networks need to be supported changes over time (Imperial et al. 2016). However, much current knowledge about governance networks tends to rely on static representations of who is involved and how they relate. By studying how an effective network evolves and sustains itself, we can learn how to govern networks over time (Milward and Provan 2000; Provan and Kenis 2008).

Existing longitudinal research has made important strides in conceptualizing the lifecycle of governance networks. Much of this work utilizes qualitative methodologies, building rich theories of how stakeholders came together and how the network and its activities evolved over time (Weber 2009; Imperial et al. 2016 and other articles in this issue). However, measuring these same networks quantitatively enables easier comparison across networks and across governance domains and therefore more generalizable lessons about how networks evolve. Nonetheless, the few papers employing statistical network analysis are based on multiple-stage survey collection, which capture overall changes between two timepoints (Berardo and Scholz 2010; Henry and Dietz 2016), but not more fine-grained temporal nuances.

This paper observes the evolution of the governance network involved in a highly collaborative process for licensing a hydropower dam in Washington state. We use text mining techniques to observe meeting participation networks from meeting minutes for a 16-year period. This offers several key advances over existing studies of network governance, in that we are able to measure participation objectively rather than rely upon self-reported participation measures and measure fine-grained interactions over time rather than aggregated summaries. Additionally, this research studies a partially mandated network. Most existing theories of network evolution stem from grassroots settings where the network evolved organically. By studying the evolution of a mandated network, and the continuation of the network after the mandate was no longer in place, we test whether mandated networks follow similar trajectories as organic ones.

In what follows, we begin with an overview of the collaborative network studied, the Baker River hydropower relicensing in Washington State. We then draw on the network governance and collaborative governance literatures to develop a series of hypotheses about how we expect the Baker River network to change as relationships evolve and policy goals change. Next, we describe our data collection and analytical approach. We use webscraping and natural language processing to encode who attended meetings and what they did at those meetings, thereby creating a network of participation (rather than just attendance). We then use temporal exponential random graph models (TERGMs) to assess each of our hypotheses. We find that the mandated network was remarkably stable over 17 years, with few changes in who occupied central nodes and with active participants more likely to sustain their engagement over time. We conclude with broader theoretical and methodological implications of this work.

Case background: Baker River hydropower relicensing

In the United States, the Federal Energy Regulatory Commission (FERC) regulates non-federally-owned hydropower facilities, to ensure that rivers--a public resource--are not overexploited for private gain. FERC issues 30 to 50 year operating licenses to electrical and water utilities; when the license expires, the facility owner must reapply for continued authorization. This reapplication process, called “relicensing,” is a mandated coordination process, as the utility is required to seek input from resource agencies and other stakeholders as they develop their license application. However, some utilities opt to create a highly collaborative process, closely involving federal and state resource agencies, local governments, nongovernmental organizations (NGOs), tribes, and others in jointly evaluating the project and proposing a new operating regime (Ulibarri 2015a, Ulibarri 2015b).

We focus on one highly collaborative relicensing, held for the Baker River Hydropower Project (hereafter "Baker") located in northern Puget Sound in Washington state. The project's relicensing, which resulted in a new operating license in 2009, was one of the most collaborative in a study of 24 relicensings nationwide (Ulibarri 2015a). Baker provides a particularly interesting case to study network evolution because it is a hybrid of multiple collaborative approaches. The authorizing legislation (the Federal Power Act) mandates a minimum level of collaborative engagement in a relicensing, making it partially externally directed process; at the same time, the Baker utility opted to create a much more collaborative approach than required, creating a partially self-initiated collaboration (Emerson and Nabatchi 2015a). Many theories of network governance focus on bottom-up, self-initiated networks (Mandell and Keast 2008; Imperial and Koontz 2007; Provan and Kenis 2008), yet governments are increasingly mandating these approaches (Bryson et al. 2006). By understanding how a network self-organized around a mandate, we can better understand how to support effective collaboration. Additionally, the externally directed mandate to collaborate in Baker vanished after the new license was issued, yet the governance network has continued in overseeing implementation of the license, providing an opportunity to observe how collaboration might sustain itself and evolve in the absence of formal government support.

The relicensing process is comprised of several phases.[1] In the first relicensing phase, *Planning and Scoping*, participants in the relicensing process evaluate what is known about

[1] In many hydropower relicensings, each phase is driven heavily by the utility. However, in the Baker relicensing most decisions throughout the relicensing were made collaboratively by a diverse set of stakeholders. In this description, we therefore frame decision-making as inclusive throughout the process.

the hydropower project and what environmental, economic, and cultural impacts it might have. During this phase, the group also decides how to structure the relicensing process, including how to share information and when and how often to meet. In phase two, *Application and Settlement Development*, stakeholders conduct a series of technical studies to evaluate the hydropower project's impacts and use the study results to negotiate a series of proposed license requirements (captured in the settlement agreement). During *Agency Review*, FERC and other certifying regulatory agencies review the submitted settlement and all associated comments to determine whether the proposed license conforms to relevant regulations. This stage ends with issuance of the new license. Finally, during *License Implementation*, the utility and stakeholder group implement the operating requirements codified in the new license. Figure 1 shows a full timeline of the Baker relicensing, including start and end dates of each phase and other major milestones.

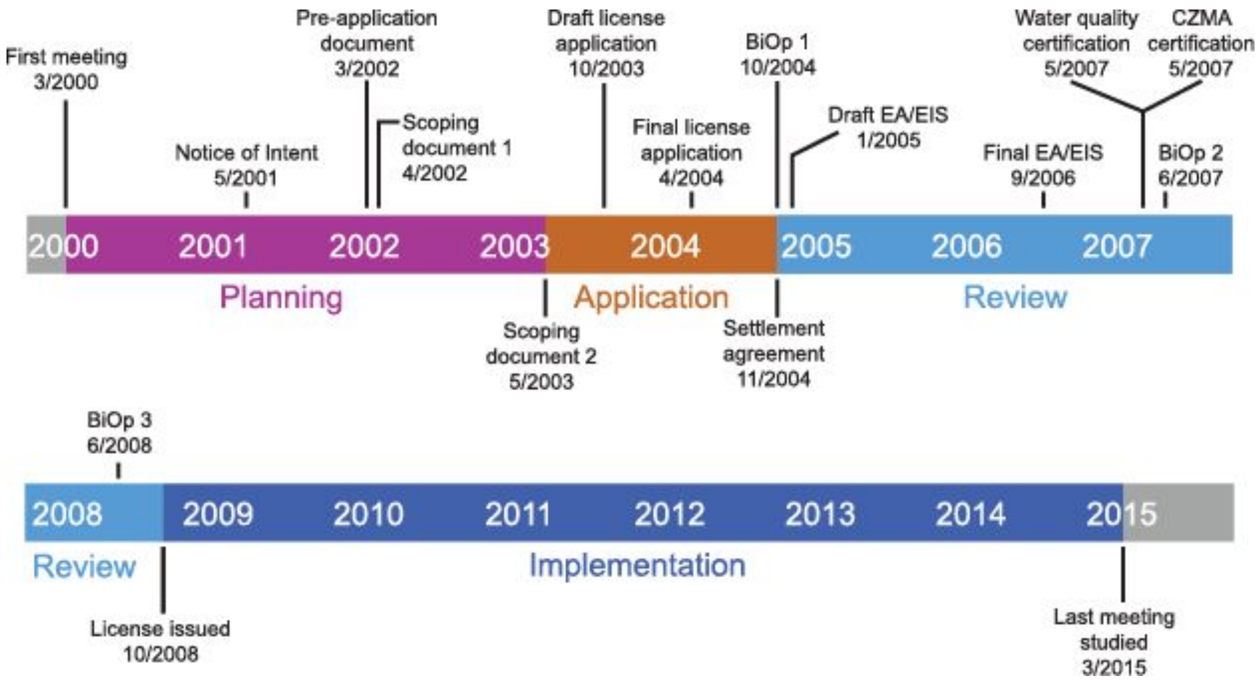

Figure 1: Process timeline for the Baker River relicensing

Theoretical framing and hypotheses

In this paper, we explore two interrelated questions. The first asks how the participation network evolved over the 16-year course of the relicensing and subsequent license implementation. In the second, we ask whether there was a discernable change in the network after the new license was issued and there was no longer any external incentive to collaborate.

While they are highly related, it's important to distinguish between *network governance*, which describes the structure of who interacts with whom in a process (Provan and Kenis 2008), and *collaborative governance*, which focuses on the interpersonal dynamics that occur in these interactions, specifically principled engagement and shared decisionmaking

(Emerson and Nabatchi 2015a). This research focuses on the network; previous research (Ulibarri 2015a; Ulibarri 2015b) established Baker to be a highly collaborative process, in which the utility engaged in deliberative, shared decision-making with diverse stakeholders throughout the relicensing. Here, we ask how the network changed over time to support that collaboration and the strong environmental outcomes associated with the new license.

Drawing on existing theories of how networks evolve over time, as well as extensive knowledge of the relicensing process (obtained through surveys, interviews, document analysis, and participant observation), we develop a series of hypotheses as to what we expect to observe in each phase of the relicensing. First, a concern in many collaborative processes, including FERC relicensing, is that differential access to sources of power skews the decision-making process from a fully egalitarian process (Purdy 2012). In FERC relicensing, the Federal Power Act (FERC's authorizing legislation) endows certain resource agencies with "mandatory conditioning authority," meaning that FERC treats their proposed license terms as requirements (rather than recommendations).[2] Thus, the playing field is inherently biased toward these agencies' opinions relative to all other participants. In qualitative observation of FERC negotiations, the utilities sometime appear to defer to these agencies more than other actors (Ulibarri 2015b), suggesting that this power dynamic could be visible in the network structure. We therefore hypothesize that

[2] Federal land management agencies (e.g., US Forest Service, Bureau of Land Management) can mandate requirements for hydropower projects on or affecting their property; the US Fish and Wildlife Service and National Marine Fisheries Service can mandate fish passage at the projects; and state water agencies can mandate requirements under section 401 of the Clean Water Act.

organizations with mandatory conditioning authority will occupy more central nodes throughout the relicensing relative to other non-utility participants.

H1: The utility and organizations with mandatory conditioning authority will occupy more central nodes and serve in leadership positions more regularly than other organizations.

Over time, governance networks become more stable and routinized (Mandell and Keast 2008). As trust and norms of interaction build, relationships become more cohesive (Imperial & Koontz 2007), leading to more stable patterns of interaction. At the same time, however, the FERC relicensing process is a time-intensive process; for Baker, we observe 210 meetings that occured during the Scoping and Settlement Development phases. Thus, organizations and individuals--who have many other potential networks and policy arenas to engage in--are only going to continue participating if they expect participating in the relicensing will serve their interests (Lubell et al. 2010). While early in the relicensing a stakeholder may have high expectations for the process, with each subsequent phase they gain more information about the costs of participating versus expected benefits. For example, once the settlement agreement has been submitted, a stakeholder has a clear signal whether their resource goals are likely to be in the license. We should expect that attendees who do not feel that the benefits of participation outweigh the costs will leave the relicensing over time.

Who are these participants that are likely to drop off? One potential indicator is the regularity with which a participant attends and participates. An actor who knows early in the relicensing whether it's worth participating (e.g., the mandatory conditioning agencies) is likely to attend regularly and vocalize their needs and interests during the meetings. These participants will occupy more central positions in the network of participants. In contrast, participants who are uncertain about the benefits may either attend less frequently or be less vocal during meetings. While some of these individuals will choose to stay, others will choose to leave as they learn more about the process during the outcomes. This indicates that we should see more peripheral actors--those who are not participating actively in meetings--choosing to leave the relicensing over time.

H2: Peripheral attendees will be more likely to drop off with each subsequent phase of the relicensing.

An additional indicator that an organization is likely to benefit from participation comes from their ability to engage in the technically and regulatorily complex process. Because all operating requirements must be based on rigorous science, engaging in the relicensing entails engaging with technical studies on the projects' hydrological, geomorphological, biological, and economic impacts. FERC relicensing is also a highly bureaucratic process, with numerous rules and statutes guiding who has the authority to do what and when. Thus, we expect organizations who have less technical expertise or resources to access that technical expertise (e.g., consultants) to leave the network at a higher rate than organizations with better access to knowledge resources.

H2a: Peripheral actors who drop out will represent organizations with less resources and/or less technical expertise.

Just as organizations evolve and grow through different stages of a lifecycle (Cameron and Whetten 1981), networks of organizations are theorized to move through particular stages in which particular organizing features and challenges are predominant (Imperial and Koontz 2007). The literature posits that self-initiated, organically arising networks--which dominate existing theory--begin in a somewhat creative, innovative, and turbulent stage of formation (Imperial et al. 2016). During this stage, actors are establishing new relationships; the trust and norms established in these relationships is key for sustaining an effective network (Mandell and Keast 2008). Additionally, because the network is evolving organically, it may not be immediately clear who relevant stakeholders are. While the convening organization may invite certain groups to participate, those groups may choose not to participate while others may drop in, leading to the somewhat turbulent nature of the early network (Imperial et al. 2016).

In FERC relicensing, formation overlaps with the Planning and Scoping phase. At this point, the relicensing group is coming together for the first time, learning about the relicensing process and the hydropower project, and learning about one another. In terms of network characteristics, this suggests that--relative to later stages in the relicensing--the network will consist of many weak ties, and that there will be high turnover in established ties as participants join or leave the group.

H3: During the planning and scoping phase, the relicensing network will be more dynamic than in later phases, with frequent changes in attending individuals and organizations.

As discussed above, many theories of network evolution posit that networks will stabilize over time (Provan and Kenis 2008; Imperial and Koontz 2007). Thus, if theories developed for grassroots networks hold in this case, we should expect to see more regularized patterns of meetings and attendance in the final stage (license implementation). At this point, the relicensing participants have less uncertainty about the process, and therefore have a better grasp of whether their interests are potentially at stake going forward (and moreover, have perhaps realized their aims already). Due to regularization of relationships and roles (Imperial et al. 2016) and a constraining of actors given the more targeted goals of license implementation, the network should be most stable during the final stage.

What makes the Baker case particularly interesting is that during license implementation, the external directive to collaborate from FERC is no longer in play. Thus, if FERC's is all that is sustaining collaboration, we would expect network patterns to become *less* stable. However, because Baker chose to use a highly collaborative approach (far beyond the policy mandate), the external directive should not matter much. Instead, theories of network evolution should hold, and the network should be especially stable during license implementation.

H4: During the license implementation phase, network patterns will become more stable.

Additional hypotheses relate to way power dynamics and leadership roles within the network are expected to change over time. Leadership plays a critical role in enabling collaborative, networked processes to function (Emerson and Nabatchi 2015a; Ansell and Gash 2008; Bryson et al. 2006). The agencies who convene the process shape the table around which decision will be made, structuring who is invited to participate and the nature of interactions (e.g., whether a network will function as collaboration or consultation) (Scott and Thomas 2016). However, just as we expect the network to evolve over time, the actors who occupy more central leadership positions should change if there is in fact a shift toward less hierarchical engagement. In FERC relicensing, the utility begins the process with the most knowledge of the hydropower facility and of the relicensing process. Therefore, they should occupy a very central role during the planning and scoping phase. Over time, however, the other participants are building relationships and gaining knowledge about the project. With that knowledge, they should begin to direct the network's conversations and decisions during later stages of the relicensing.

H5: The utility will serve as a leader a higher proportion of the time during the planning and scoping phase than in later phases.

Analytic Framework

Temporal network models

Each hypothesis advanced above describes an expectation regarding interactions between relicensing process participants. Thus, our basic units of analysis are interactions between participants observed within a given time period. In the following section concerning field methods and data collection, we describe how these observations are generated. This section describes the analytic approach we use to test our hypotheses.

The most basic approach for modeling such interactions would be to fit a logistic regression model that estimates the probability of an interaction between actor A and actor B as a function of relevant actor attributes and grouping variables such as organizational affiliation or time period. However, from a broader perspective these individual connections can be understood as constituting a network. This network is comprised of a "node set" where each node is an actor involved in the relicensing process, and edges, or ties, defined by observed interactions between two actors. Network analysis requires a slightly different statistical approach because dyadic observations (e.g., presence or absence of a tie between actors A and B) are expected to exhibit hyperdyadic dependency (Cranmer and Desmarais 2016). This dependence means that "the likelihood of a tie may not only be a function of individual characteristics of actors who share the tie, but also a function of the presence or absence of other network ties" (Koskinen and Daraganova 2013, 51). Accordingly, we assume that the data generating process a function not just of the attributes of the focal dyad (e.g., attributes of actors A and B in this case) but also of the values of other dyads and broader network structure. Indeed, a host of research analyzing policy networks reveals how patterns of bridging, bonding, reciprocity, and mutual association amongst multiple network actors influences the value of any one network tie (e.g., Desmarais and Cranmer 2012a; Berardo and Scholz 2010; Berardo 2014; Lubell et al. 2014; Adam D. Henry et al. 2011; Scott 2016).

Exponential random graph models (ERGMs) are an increasingly prominent tool used to assess relational patterns amongst policy actors subject to such complexities. Because ERGMs are well-established at this point as a tool of inference in policy research, we refer the reader to prior applications within the extant literature (Lubell et al. 2012; Robins et al. 2012; Desmarais and Cranmer 2012a; Jasny 2012; Park and Rethemeyer 2014; Ingold and Leifeld 2016) rather than revisit the basic method in greater detail. In brief, an ERGM produces output similar to that of a logit model, with coefficient estimates that reflect the conditional log-odds of observing a given network tie. However, an ERGM sets the entire network as the dependent variable, and then compares the observed network against a simulated distribution of hypothetical networks that are generally similar to the observed network, serving to identify network patterns that are unlikely to be random net of other features (e.g., if there are more triangle structures than observed in other networks with a similar number of total edges). One important caveat to emphasize is that ERGMs do not assume complete interdependence, but rather simply make no assumptions about dyadic independence (Cranmer and Desmarais 2016). Thus, ERGMs are a way to make more accurate probabilistic assessments of network ties by allowing for, rather than assuming away, dyadic interdependencies. If hyperdyadic dependence is not present, and ERGM reduces to a logistic regression model (Cranmer and Desmarais 2011).

Most of the extant literature has employed cross-sectional ERGMs; since we observe repeated interactions over the course of many years, and are explicitly interested in how ties arise and dissolve over time, a longitudinal specification is necessary. Temporal ERGMs, or TERGMs, are relatively uncommon in empirical applications, in part because collecting repeated network measures can be difficult, particularly for the survey-generated measures used for most ERGM analysis in the policy and management literature. Thus, it is instructive to describe the basic mechanics of TERGMs. To fit TERGMs for our analysis, we use a bootstrapped pseudo-likelihood estimation method developed by Leifeld et al. (2015) and included in the *xergm* package (Leifeld, Cranmer, Desmarais, et al. 2015) in R (R Core Team 2016). Thus, in what follows we refer to the bootstrapped TERGM, or BTERGM, model.

BTERGMs expand upon the basic ERGM framework by modeling a longitudinal sequence of discrete time network observations. A network at time t follows an ERGM distribution in which the vector of model parameters *v* is a function of the observed network at time *i* (G^t), as well as the network at q preceding time points (Desmarais and Cranmer 2012b). Thus, model terms can reflect both current dependencies and intertemporal dependencies such as edge stability or the sequential formation of mutual ties or triangle structures. This longitudinal perspective is theoretically important, because the very idea of structural dependency implies that individual network ties are generated conditional on the rest of the network, meaning that the network is expected to undergo sequential updating (Desmarais and Cranmer 2012a).

We observe the relicensing process (both pre- and post- license reissuance) over the course of 16 years. During this time, more than 600 meetings were held (this is described in more detail in the data section below). For every observed meeting, we catalogue attendees, their organizational affiliation, and organization type. We follow the methodology of Ulibarri and Scott (2017) for coding network ties on the basis of recorded actions by attendees, such as making a presentation or participating in a group discussion. For a given meeting, if (for example) an attendee gives a presentation, this individual is coded as having a tie value of 1 extending to every other meeting attendee. For instance, if actors A and B both show up at the meeting, and actor A gives a presentation, $y_{A \rightarrow B} = 1$, while if actor B does not say anything, $y_{B \rightarrow A} = 0$. These ties are directed ties, meaning that the tie extends unidirectionally from the presenting or speaking actor to assumed listeners, and that more generally that the value of a tie from actor A to B (y_{AB}) does need not equal the value of the reciprocal tie from actor B to actor A (y_{BA}). These observations provide direct evidence of communication ties between participants.

Having coded behavioral ties occurring at each meeting, we then group observations by six-month periods. The end result is a series of 32 semi-yearly network observations, where all ties observed within a given year are combined to constitute the observed network for that period. We then model changes in the relicensing network across these 32 periods. Our rationale for grouping the data in this way are threefold. First, while one could analyze this process on a meeting-by-meeting basis (i.e., conceptualizing each meeting held during the relicensing process as representing one manifestation of the collaborative network), individual meetings are held by subgroups that concern specific topical components of the overall relicensing process (e.g., aquatic resources, or recreation). Two temporally adjacent meetings might concern two completely different subjects, and thereby involve different network subgroups; it would be incorrect to assume on this basis that the network has changed!

Second, we also rule out treating each topical workgroup as a separate longitudinal network (e.g., modeling the aquatic resources work group meetings as a separate network than the terrestrial resources work group meetings). Considerable overlap in attendance is observed across topical groups, and each meeting was open to all network actors in any case. Moreover, many joint meetings between subgroups were held over the course of the process, indicating that participants themselves did not view these topical areas as comprising separate networks either. All of these factors make it difficult to justify not treating all subgroups as components of a holistic network.

Third, from an analytical perspective, since we observe many collaborative actions over the course of each year, but only a relatively small number of individuals attend any given meeting, a meeting-by-meeting focus presents an overly sparse portrayal of process interactions. While each meeting reveals collaborative actions, no meeting should be viewed as representing the entirety of the relicensing process at any one time period. Thus, we believe it is more conceptually appropriate to treat the network on biannual basis, from May to November to May; this provides a relatively large number of time periods within each phase, while still providing enough breadth to capture several meetings within each period. Having discussed the BTERGM model, we now turn to model specification, specifically how we test our hypotheses using a combination of endogenous structural parameters and exogenous covariates.

Explanatory variables

Our first hypothesis, H1, anticipates that organizations with the authority to condition license requirements will be central network actors. This implies that these actors will have, on average, more outgoing ties to other actors. To test this, we fit a categorical attribute designating actors affiliated with organizations that have mandated authority with the process that specifically models only outgoing edges, i.e., the "sender" of a tie. This term (*nodeofactor* in the *ergm* package) adds a statistic reflecting the number of times a designated node (in this case, one with mandatory authority) interacts with another stakeholder (Handcock et al. 2014). We fit separate indicators for utility-affiliated actors and for actors affiliated with mandatory conditioning authority (USFS, WDOE, and NMFS).

Hypothesis 2 (H2) anticipates that participants who are peripherally involved during early process phases will be more likely to stop participating as time goes on. While there are many possible ways to operationalize the concept of peripherality, the general concept is that core actors are those amongst whom there are "relatively strong, direct, intense, frequent, or positive ties" (Wasserman and Faust 1994, 8:249). One way to patterns of interaction in terms of core versus peripheral involvement is via Google's PageRank algorithm (Brin and Page 1998; Friedkin and Johnsen 2014). The algorithm, which we operationalize using the *igraph* R packages (Csardi and Nepusz 2006) uses an eigenvector strategy to assess network centrality, but in doing so normalizes for the number of ties possessed by a given actor and weights links from other central actors more heavily. A higher PageRank represents a more central actor, and a lower PageRank a peripheral actor. For each observation period, we compute each actor's PageRank centrality score and then fit each actor's core rank in the prior period as a model covariate. For a given network observation, the total PageRank scores for all nodes sums to 1; because this results in very small values, we scale these scores by a factor of 100.

H2A further expects that peripheral actors from low-resource organizations, and/or peripheral actors who are non-technical in orientation, will be most likely to drop out over time. To test this, we further fit an additional indicator for individuals who are affiliated with the utility, federal agencies, state agencies, and various technical consulting firms. We then interact the indicator for these three types ("technical actors") with an actor's centrality score from the prior year in order to compare dropout across these two different types of actors. The interaction terms results in separate slope adjustments for prior core status by technical and non-technical actors.

Next, H3 addresses the question of tie dynamics over the course of the process. Dyadic stability refers to how likely it is that a tie present at time *t* will also be observed at time t+1, or conversely that a tie not present at time *t* will also not be present at time t+1. To assess dyadic stability, we incorporate a memory term that tests for how likely it is that a dyad changes in value tie (i.e., $y_{ijt} = 0$ and $y_{ij(t+1)}= 1$, or $y_{ijt} = 1$ and $y_{ij(t+1)}= 0$) between observation periods. H1 anticipates a contrast in tie dyadic stability between different phases of the relicensing process. By interacting a dyadic stability term with process phase, we are able to estimate a unique measure of dyadic stability within each phase.[3]

H4 poses that network patterns will become increasingly stable during the implementation phase. Whereas H1 focuses both on tie creation and dissolution during initial process

[3] While this is fairly simple in principle, in practice there are a series of steps required to implement such an interaction term for use with the BTERGM package. In brief, we first create a matrix for each time period 2 to t (where t is the final time period), in which each matrix is filled with values of 1 (if a given tie was present at time t - 1) and -1 (if a tie was not present). This list of matrices are fit as an edge covariate using the *edgecov* term, which simply serves to model how the previous dyad value predicts the current dyad value. The interaction effect is then incorporated using the *timecov* term to interact the list of prior tie values at each period with process phase (i.e., this adds a second term that adjusts predicted stability, only for periods that occur after the planning and scoping phase). We are immensely grateful to the BTERGM package author, Dr. Philip Leifeld, for his guidance in determining how to specify models of interest.

phase, H3 poses a slightly different expectation: that edges will be more stable during the implementation phase. Thus, rather than examining the collective tendency both for ties that did not already exist to form and for ties that previously existed to dissolve, we fit a memory term that captures edge autoregression. This term focuses specifically on the likelihood that an edge observed at time t will also be observed at time t+1. We further interact this term with an indicator for the project phase. This results in separate tie stability estimates for each phase. The model elements used to test H3 and H4 work quite similarly since both relate to the autoregressive tendencies of tie across periods. As we discuss in more detail below, in order to avoid issues related to multicollinearity, we fit separate models that omit either the terms used to test H3 or those used to test H4.

Table 1: Model specifications used to test hypotheses

Theoretical hypothesis	*Operationalization*
H1: Utility and mandatory authority orgs will be central nodes.	Predicted difference in outdegree value for utility and for mandatory authority organizations
H2: Peripheral actors are more likely to drop out as process proceeds.	PageRank centrality score in prior period
H2a: Peripheral actors who drop out will be from low-resource orgs or have less technical expertise.	PageRank centrality score in prior period interacted with organization type
H3: During the planning and scoping phase, the relicensing network will be more dynamic than in later phases, with frequent changes in tie values across periods.	Probability of tie change (0→1 or 1→0) between t and t+1, interacted with process phase
H4: Network patterns will become more stable during implementation phase.	Probability of that tie equals 1 at time t + 1 given that tie equaled 1 at time t, interacted with process

	phase
H5: Utility-affiliated actors will be the leader more during planning/scoping than in later phases.	Interaction of indicator for utility organization with time covariate

Finally, H5 holds that actors affiliated with the licensed utility are more likely to play a leading role early on in the process because they enter with the most knowledge regarding the project itself and the broader resource context. In order to test this, we interact an indicator for utility affiliation for outgoing ties with a time covariate; this essentially serves to fit a differential temporal effect for outgoing ties (i.e., participation actions) by actors affiliated with the relicensing utility. We can then compare this interaction term with the basic time effect to gauge the extent to which leadership by utility actors differs from other participants.

Control variables

In addition to the terms described in table 1, which relate directly to hypotheses of interest, there are several other terms included in each model. First, because meeting attendance is a prerequisite for having any interaction ties, it is important to control for the number of meetings each actor attends within a period. Along with this exogenous term, we include a set of endogenous structural terms that capture baseline network processes; these “sufficient statistics” (Levy et al. 2016) are calculated at the network level and reflect the overall pattern of ties within the network. First, an edges term, which is analogous to an intercept term, is used to establish the baseline probability of an edge (in other words, this term reflects the probability of a tie between two randomly selected nodes, all else held constant). Networks also frequently exhibit patterns of reciprocity, in which a tie from node j to node i is much more likely given that tie already exists from i to j. Thus, we fit a term ("mutual") that estimates the probability of a reciprocated tie.

Each node in a network also has a degree value, which refers to the number of edges incident on that node. For instance, an actor who has four incoming ties within a given time period has an in-degree value of 4, and an actor who has 10 outgoing ties has an out-degree value of 10. The degree distribution across network actors reveals the extent to which a network is more centralized, with ties tending to cluster on popular actors, or dispersed, with a less skewed degree distribution. A classic approach to modeling network degree is to fit star terms, for instance 2-stars, which count the number of instances where two edges are incident on the same actor (e.g., A→ B and C → B). However, controlling directly for specific star structure (e.g., 2-stars, 3-stars) can pose computational problems because of the interrelation between density and star counts (Snijders et al. 2006). A preferred technique for accounting for degree is to use a geometrically weighted degree (GWD) statistic, which models the distribution of degrees across all nodes as a network-level attribute (Hunter and Handcock 2006).

A GWD statistic is a weighted count of degree values, which in this case represents the activity spread across process participants. Incorporating the GWD statistic into an ERGM results in a coefficient that reflects the extent to which a network has a higher or lower degree distribution variance (we return to this interpretation in the results section below) (Hunter 2007). This result is partially determined by the shape parameter θ_s used to conduct geometric weighting; the shape parameter determines how much less adding an additional interaction to an already highly active participant contributes to the overall GWD statistic relative to adding a tie to a less active participant (Levy et al. 2016). Because this network consists of directed interactions, each node has an indegree (the number of observed interactions from other actors to the focal node) and an outdegree (the number of observed interactions by the focal node). We fit a GWD term for indegree (GWID).[4][5]

Finally, we also fit a geometrically weighted edgewise shared partner (GWESP) statistic. The GWESP specific functions similarly to the GWD statistic, but serves to account for the network transitivity. Transitivity is the tendency for triadic closure, for instance the increasing likelihood of an interaction occurring between A to C if both A and C already interact with a third actor, B. The GWESP statistic down-weights each additional “shared partner” that two nodes have, again down-weighted using a fixed shaped parameter. This shape parameter serves to determine the relative contribution of adding one more shared

[4] An GWD outdegree term was tried within various model specifications, but was not shown to account for a meaningful amount of variance net of other model terms.

[5] For the GWID shape parameter, we use a relatively high shape parameter of 2. A higher shape parameter means that the GWID statistic is more responsive to changes in high-degree nodes (Levy 2016). The way in which we code interactions results in high in- and out-degree values. Since a meeting attendee is assigned an ingoing tie from active meeting participants, the change, for instance, from an indegree value of 1 to an indegree value of 2 is largely inconsequential, because this would mean that only two individuals actively participated across all of the meetings attended by that actor. Thus, relevant changes in degree are changes in higher degree nodes, because these changes more accurately reflect observed dynamics. That said, it is also possible that the most important distinction is between active and passive participants, and thus that the most important degree change is going from an out-degree of 0 to 1 in a given period. We revisit this issue in the context of the results below.

partner, depending upon the number of existing shared partners (Levy et al. 2016). Intuitively, the GWESP term is valuable because one would not expect each additional shared partner to have the same transitive pull. In terms of estimation, Levy et al. (2016) further note that a GWESP parameter should generally be included in conjunction with GWD terms to avoid potential confounding between network centralization (concentration of ties on certain nodes) and clustering (subgroups with dense tie patterns). One terminology issue to note is that our network consists not of partnership ties but of observed patterns of interaction; nonetheless transitivity is key driver of network patterns in general (Lusher et al. 2013), and the GWESP term is useful for accounting for transitive patterns in stakeholder interactions without controlling for the counts of multiple specific structures.[6]

Data and Field Methods

Data Sources

The primary source of data for this analysis are publicly available documents available through Puget Sound Energy's (PSE) website. PSE maintains an online repository of documents related to the Baker River relicensing process and ongoing management activities, including meeting agendas, meeting summaries, and technical reports. For this

[6] As with the GWID term, we use a relatively high shape parameter of 2 in order to emphasize changes in higher shared partner counts. Note that because the value of the shape parameter is part of what determines the coefficient value, for robustness we conducted sensitivity tests, varying both fixed shape parameters between 0.01 and 3 and refitting the baseline model presented below. This is discussed further in the results section.

analysis, we draw from summaries published for 591 different meetings held between years 2000 and 2015. We also identify 48 meetings for which no summaries are available. These unobserved meetings were held by a variety of workgroups over the course of several years; thus, we are confident that these missing data do not bias our analysis. To ensure reproducibility in the case that PSE's storage practices change, the authors have also archived every document. These are available upon request.

Text Analysis

Observations are extracted using natural language processing, a machine learning method described in the field methods section below. In total, we observe 774 unique actors attending at least one of 591 different meetings.[7] In brief, each time a meeting summary document references an action taken by an attendee (e.g., "Dwayne Carter suggested that the plan be revised."), this observation is categorized in terms of the attendee and the type of recorded action. In order to avoid recording names that are not actual meeting attendees (e.g., the authors of a study or report cited at a meeting), we cross-reference the set of extracted entities against a carefully cleaned meeting attendance database, and filter out all extracted actions that do not pertain to meeting attendance behaviors. Using the VerbNet lexicon[8] (Schuler 2005), we then are able to sort observations by the type of verb. This

---

[7] Note that we are aware of 639 meetings that were held, but 48 either had no documentation or were only documented by a pre-meeting agenda document that did not contain any record of attendance or participation.

[8] VerbNet is an index of verbs from the English language that structures verbs into a hierarchical tree of use classes which can be roughly understood as general groups of actions, roles in sentence structure, and frames. While the granularity of VerbNet allows one to be either very specific or very general in assigning verbs to categories, we rely on verb classes only— this results in 271 possible classes for verbs. We further limit these classes via semantic coherence resulting in a much more parsimonious set of classes. For example, "remove",

facilitates going beyond simple tagging of actions to a more fine grained description of participation. The VerbNet lexicon identifies verbs that are associated with predicative complements, for instance a sentence such as "Kimberly Jones will report on the team's progress next week." These types of sentences are evidence not of actual meeting behaviors, but rather of planned future actions, which should not be counted. Thus, we drop all observations corresponding to verbs with predicative complements.

It is also important to acknowledge these data should be considered somewhat "noisy" in that there are potential false positives (attendance and/or participation actions incorrectly assigned to a given stakeholder) and false negatives (attendance and/or participation that our coding mechanism fails to correctly capture). In part, this is a tradeoff inherent with respect to the use of automated text analysis; while automation allows us to capture the full breadth of the relicensing process, there meeting summaries contain complex language structuctures, formatting irregularities, and other idiosyncrasies that are difficult to account for. That said, the raw data also exhibit ambiguity that hand coding cannot necessarily solve; for example, there are instances in the meeting summary documents where a pronoun or a collective noun is used that cannot be disambiguated to attribute the action to a specific, named individual. We are confident, however, that these sources of noise do not inhibit the usefulness of these data or unduly bias our results. First, since we are able to draw upon data from almost 600 meetings, the breadth of these data should enable accurate modeling of the stakeholder network. Moreover, there is no reason to

"banish" and "clear" are each "Verbs of Removing". A table of these classifications is available at http://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf.

anticipate that particular stakeholder are more or less likely to register as false positives or false negatives; thus, estimates are unbiased in this respect even if subject to stochasticity. Finally, the coding strategy outlined in the prior section, which aggregates ties by six month periods, serves to attenuate coding errors from any one meeting, since the stakeholder network for a given six month period is based upon many different meetings that occurring during that time. To more fully explore the implications of our automated coding strategy relative to a hand coding-based approach, appendix A compares automatically generated participation data with hand coded meeting participation data from a random subset of 49 meeting summaries. Overall, automated coding records higher attendance (averaging 2.96 more attendees across the 49 meeting subsample) and a higher number of actions taken by unique stakeholders (averaging 0.38 more unique observed participants across the 49 meeting subsample). As compared to the average number of attendees and unique participants in the hand-coded subsample, this means that the automated coding strategy overestimates the number of participants per meeting by around 21%, and underestimates the number of actively participating attendees by around 6%. Thus, machine coding is expected to prove conservative with respect to the breadth of participation (by overlooking some actions taken in meetings) and slightly too liberal with respect to attendance (by incorrectly coding some actors as being in attendance). We assume that particular types of stakeholders are not necessarily more or less likely to be overlooked by machine-coding. Thus, these sources of error are expected to influence network-wide statistics (e.g., density), but not bias structural comparisons between particular types of actors or between different time periods. We discuss this issue more fully below in the context of the results.

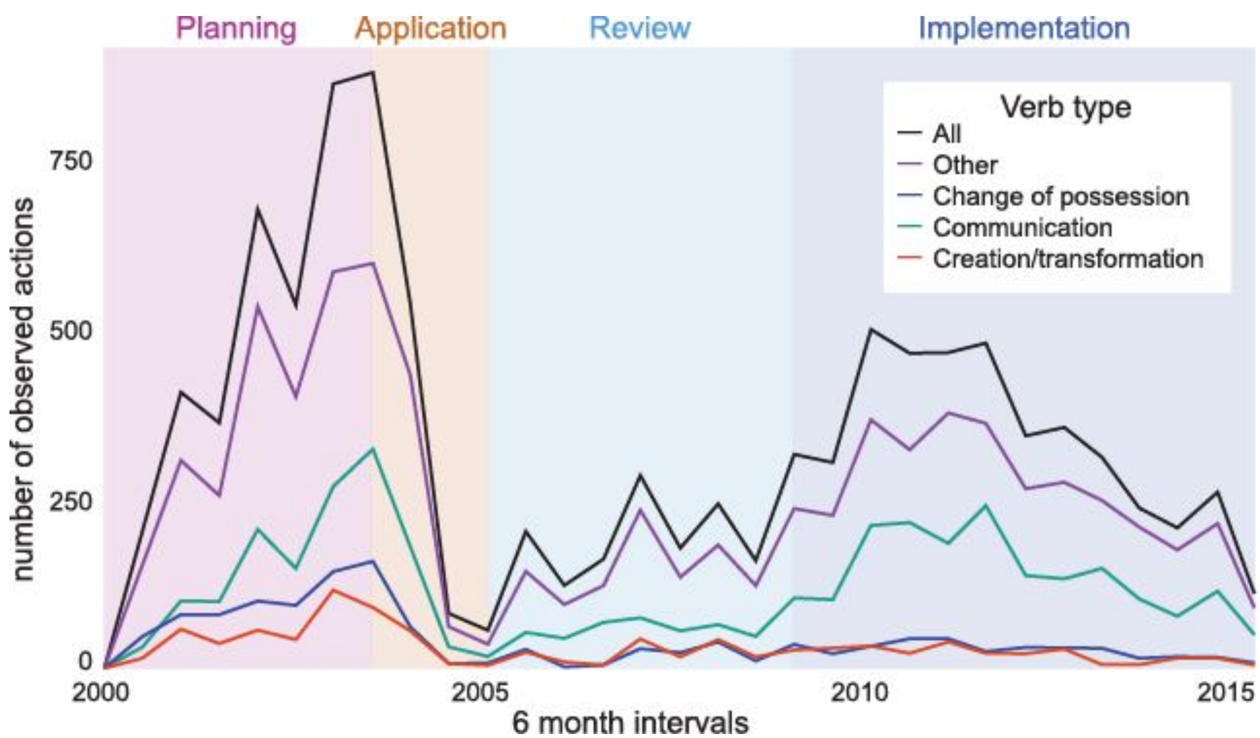

Figure 2: Number of stakeholder engagement actions observed by period and verb type

Further, meeting attendees can participate in different ways, such as giving a presentation or asking a question. We use the VerbNet lexicon as a way to categorize different types of interactions. Figure 2 shows the frequency of all observed interactions, as well as the three most commonly observed verb types (communication verbs such as "report", "explain", or "suggest", change of possession verbs such as "provide", "present", or "give"[9], and

[9] Which in this case clearly represent information exchange as opposed to the exchange of physical goods.

creation/transformation verbs such as “draft”, “work” , or “develop”) and an “other” category comprised of all but the three aforementioned verb types. Observations are further grouped six month interval and process phase.

In terms of measuring the the evolution of this collaborative stakeholder network over time, it is not necessarily clear whether some types of engagement should be emphasized or excluded. One weakness of focusing on specific verb types is that the text analysis tools we use struggle to account for idioms and colloquial speech found in the meeting minutes. A prominent example is that there are many instances in meeting summaries where a stakeholder is reported to have “walked” the group through a report, comments, or issue. The VerbNet lexicon codes the de-lemmatized version of this verb, “walk”, as a “Verb of Motion,” which at face value would seem to a non-relevant observation that should be dropped (since meeting activities are not really kinetic in nature). However, a contextualized reading reveals that “walk” in this case is an expression that in fact refers to a stakeholder having led a discussion or given a presentation, which are activities of direct interest for this analysis! Thus, focusing on a specific set of identified verb types risks excluding observations in a non-random way to the extent that certain notetakers or process subgroups use different nomenclature in their documentation.

Results

Network models can be interpreted in two different ways. The first is a global perspective that considers the probability of an observed network over the distribution of networks that could be observed (Heaney and Leifeld 2016): a positive parameter estimate means that a tie that increases the value of the corresponding statistic (e.g., the number of mutual tie patterns observed in the network) is more likely than one that does not. The second is a micro-level interpretation that focuses on the probability of observing a particular network tie as a co-determined by endogenous network structures and exogenous covariates (Desmarais and Cranmer 2012a).

Structural terms such as *mutual* are endogenous to the model because they incorporate measurements on other ties in the network. For the *mutual* term, this means the predicted change in the odds of observing a tie from actor *i* to *j* in time *t* given that a tie from actor *j* to *i* exists in time *t* as well.[10] In the case of exogenous variables such as whether an actor is affiliated with the licensed utility, the coefficient is the predicted difference in the probability of observing a tie from actor i to actor j given that actor i is from represents the utility. The coefficients associated with exogenous covariates or endogenous structural term alike act marginally upon the log odds that $Y_{ijt} = 1$. Further, we also incorporate a cubic time trend meant to account for overall change in interactions across the observation period.

[10] Desmarais and Cranmer (2012a) note that this conditional estimation process for ERGMs is consistent with the expected data generating process. A network is expected to evolve over time via a sequential updating process (i.e., ties are formed sequentially, rather than all at once, in part based upon the status of existing ties). Because an ERGM models network structures as conditional upon surrounding structures, it appropriately describes this conditional, sequential updating process.

Because the hypotheses posed above advance several time-related questions, we use a series of models rather than fit all terms of interest simultaneously. The model specification shown in figure 3 includes a series of endogenous structural terms that account for general patterns of interaction, a control for number of meetings attended, indicators for whether a given individual is affiliated with the licensed utility or an organization with mandatory signing authority, and a polynomial time trend meant to control for overall change in interactions across time. This model is used to examine H1, that actors affiliated with the licensed utility and mandatory authority agencies will be more central actors on average. Each bar in figure 3 represents a bootstrapped 95% confidence interval (see (Leifeld, Cranmer, and Desmarais 2015)) estimated for a given coefficient. Complete tabular results for all models presented below are provided in appendix B. All models presented herein use the same fixed shape parameters for the geometrically weighted network terms.[11] Note that the model in figure 3, and all subsequent models, each contain an intercept term[12] that is not shown. These intercept terms are presented in the tabular results found in appendix B, but since these terms are of a larger magnitude than other model coefficients, omitting them from the graphical results greatly improves presentation clarity.

[11] As noted in the methodology section above, the value of the shape parameter can influence coefficient estimates. Thus, we also graph estimates generated with different shape parameter values in appendix B to understand how these specification choices influence models results. Finally, goodness-of-fit diagnostics are presented in appendix C.
[12] In an ERGM, the *edges* term acts as a model intercept controlling for baseline density. Interpretation is quite simple: exp(edges) gives the baseline odds of observing a tie from actor i to actor j with all other variables set to zero (which is of course not of substantive interest in this case).

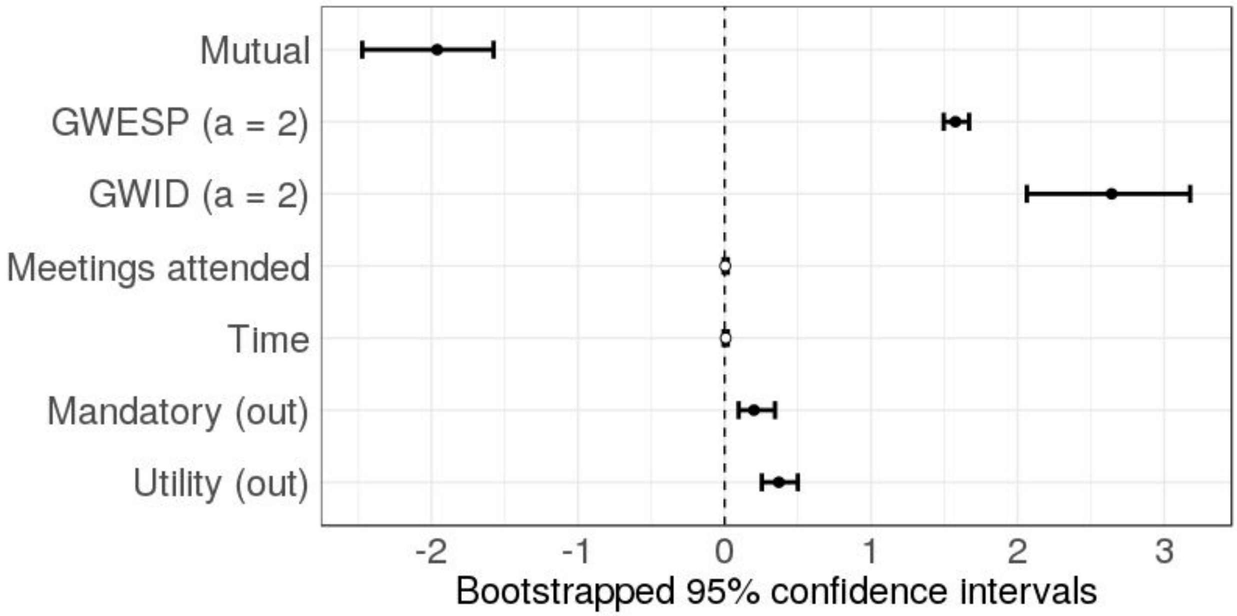

Figure 3: Coefficient estimates for model specification testing for centrality of utility and mandatory authority organizations

The two coefficients of substantive interest in figure 3 are the Utility (out) and Mandatory (out) terms; each term is interpreted as comparing the likelihood of an interaction tie extending from a utility affiliated actor or a mandatory authority organization affiliated actor, as compared to other types of participants. Figure 3 shows that all else equal, both types of actors are more likely to have an outgoing tie. Recall that outgoing ties represent instances where a given actor does something such as give a presentation to other meeting attendees. Thus, both terms indicate that representatives of utility and mandatory authority organizations occupy more central leadership roles in group meetings.

Next, H2 examines whether a peripheral actor in one period is less likely to be involved in a subsequent period. An additional specification then tests H2A by interacting centrality in the prior period with an indicator for high resource organizations. Just as with a standard logit model, interaction term coefficients do not paint a straightforward picture since the predicted marginal change is not constant across different combinations of the relevant explanatory variables (Scott Long and Freese 2006). Tabular results for the models specified to test H2 and H2A are presented in appendix B; however in order to more clearly represent the potential implications of this interaction, we follow the empirical strategy of Heaney and Leifeld (2016) and Czarna et al. (2016) and compare predicted edge probabilities at different explanatory variable combinations. Specifically, we generate predicted edge probabilities and then plot these probabilities against prior centrality rank, conditional on whether or not the actor is from a high resource organization. Further, because the impact of prior centrality rank is not substantively interesting for actors who did not attend any meetings in the prior period, in figure 4 we focus on predicted probabilities in time t for actors who attended at least one meeting in time t - 1.

Figure 4 shows that for all actors, the more central an actor is in one period, the more likely said actor is to be involved in interactions in the subsequent period. This lends support to the expectation that more peripheral actors are less likely to persist in process involvement. However, there is little observed difference with respect to the impact of prior centrality conditional upon organization type. While actors representing high resource organizations are more likely to precipitate interactions, the estimated difference between organization type remains fairly consistent regardless of prior centrality.[13]

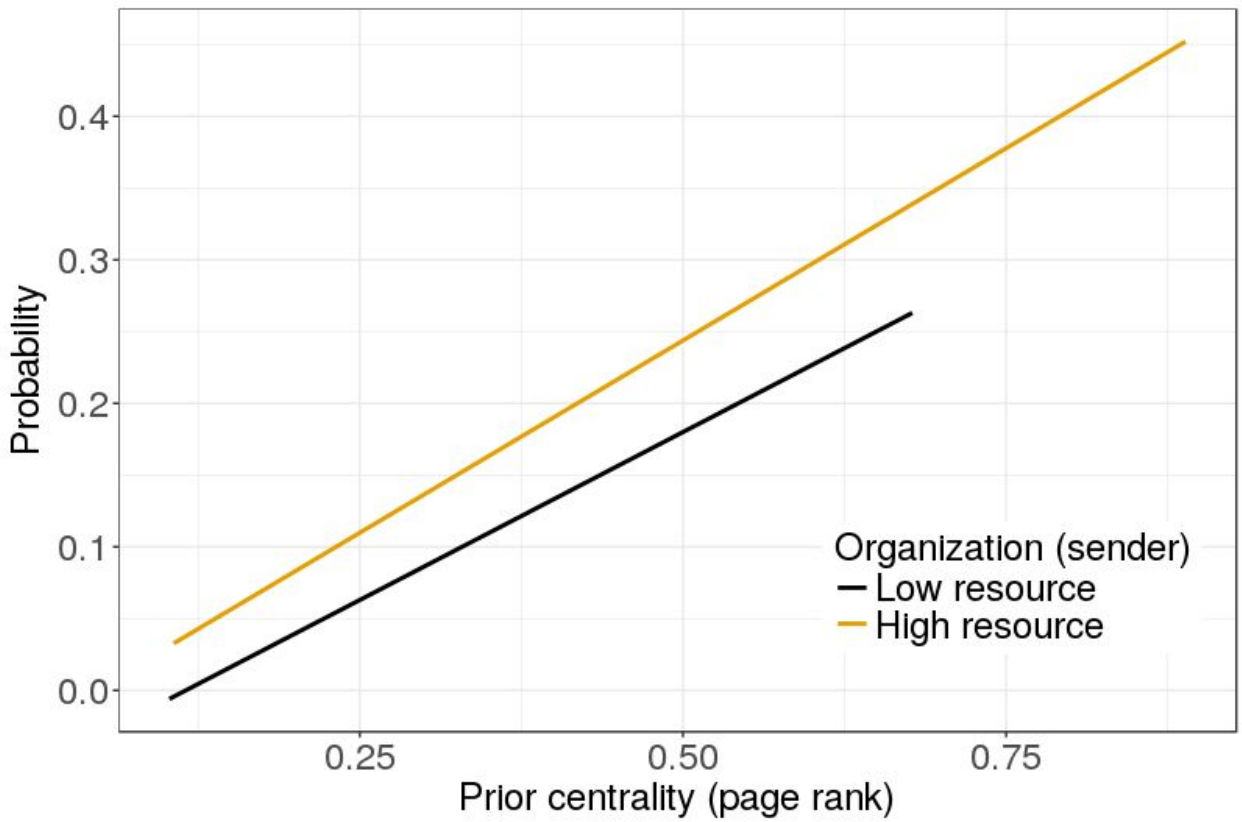

Figure 4: Predicted tie values by prior time period core rank, organizational resource status, and phase

Next, the model results shown in figure 5 contain a series of additional terms that compare dyad stability during the planning phase and in the three post-planning phases. Again, full tabular results are shown in appendix B. Each dyad stability term is a network statistic that

[13] Note that within the predicted probability line for low resource organizations cuts off before the high resource organization line due to lack of support beyond this point. In other words, only high resource actors have the highest centrality scores.

counts the number of dyads (pairs of network actors) that maintain the same tie value between time periods (i.e., present in both periods or absent in both periods) during each phase. Because the statistics for which these coefficients are estimated are calculated as the total number of dyads that remain stable between consecutive periods, a positive parameter estimate indicates that ties that existed in the prior period are more likely than ties that did not, all else equal.

The results in figure 5 do not strongly support the expectation that the initial planning/scoping phase (the first 3.5 years of the process) would exhibit less stable patterns of interaction. Overall, dyads exhibit stability throughout the process. Holding everything else constant, a dyad that had a tie in time t-1 is 158% more likely to have a tie during time t than one that did not during the planning and scoping phase (exp(0.46) = 1.58). During post-planning phases, dyads remain stable, but the predicted change in odds decrease to 144% (exp(0.46-0.09) = 1.44). As shown in figure 5, however, the confidence interval for the interaction parameter spans zero.

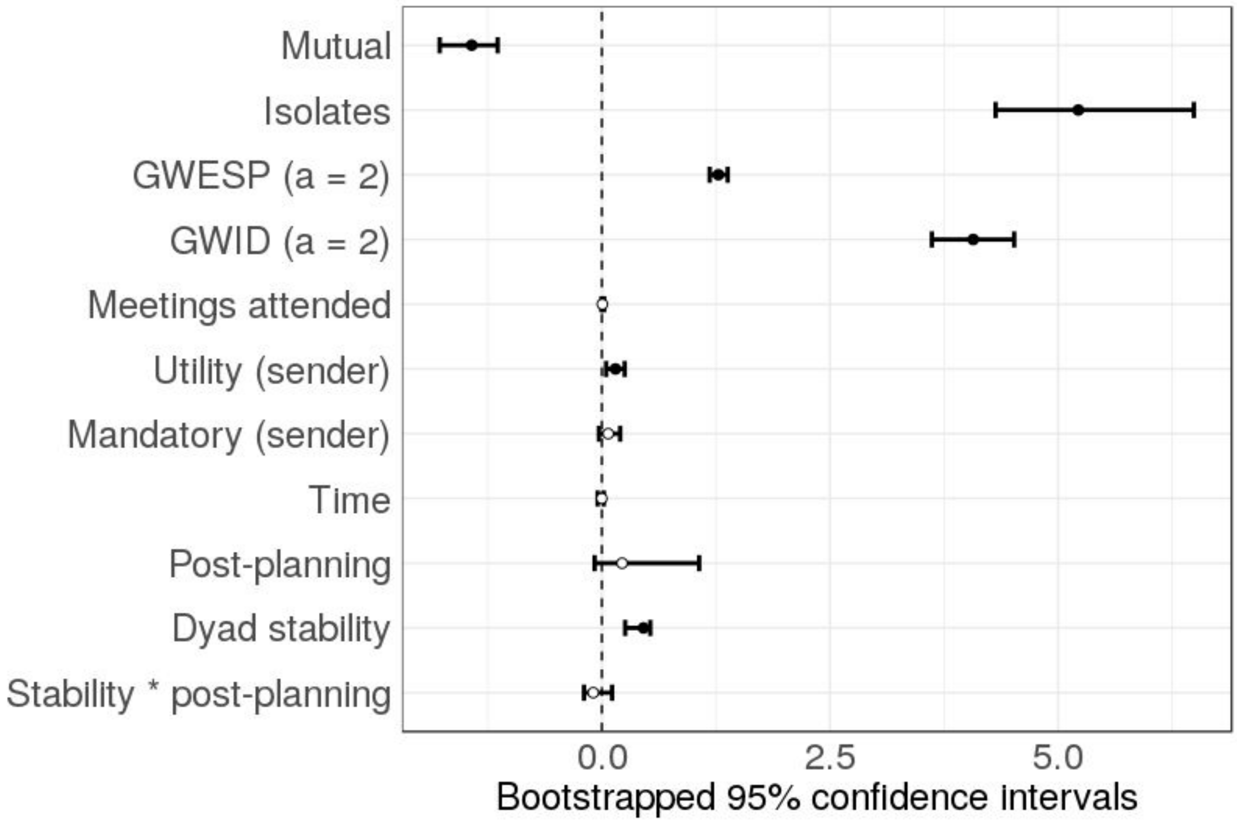

Figure 5: Coefficient estimates for model comparing tie dynamics during and after scoping

H4 also concerns tie changes between periods, but focuses on comparing tie stability pre- and post-implementation. To test this, we again fit an autoregressive term, but in this case each statistic counts the number of ties that are maintained within subsequent periods. Figure 6 shows that interactions remain relatively stable throughout the process. In pre-implementation phases, an interaction that occurred in one period is 239% more likely to recur in the following period, all else equal (exp(0.87) = 2.39). Contrary to the expectation of hypothesis 4, interaction ties are less stable during the implementation phase, as the estimated multiplicative change in odds falls to 196% (exp(0.87-0.19) = 1.96). However, the confidence interval estimated for this parameter spans zero.

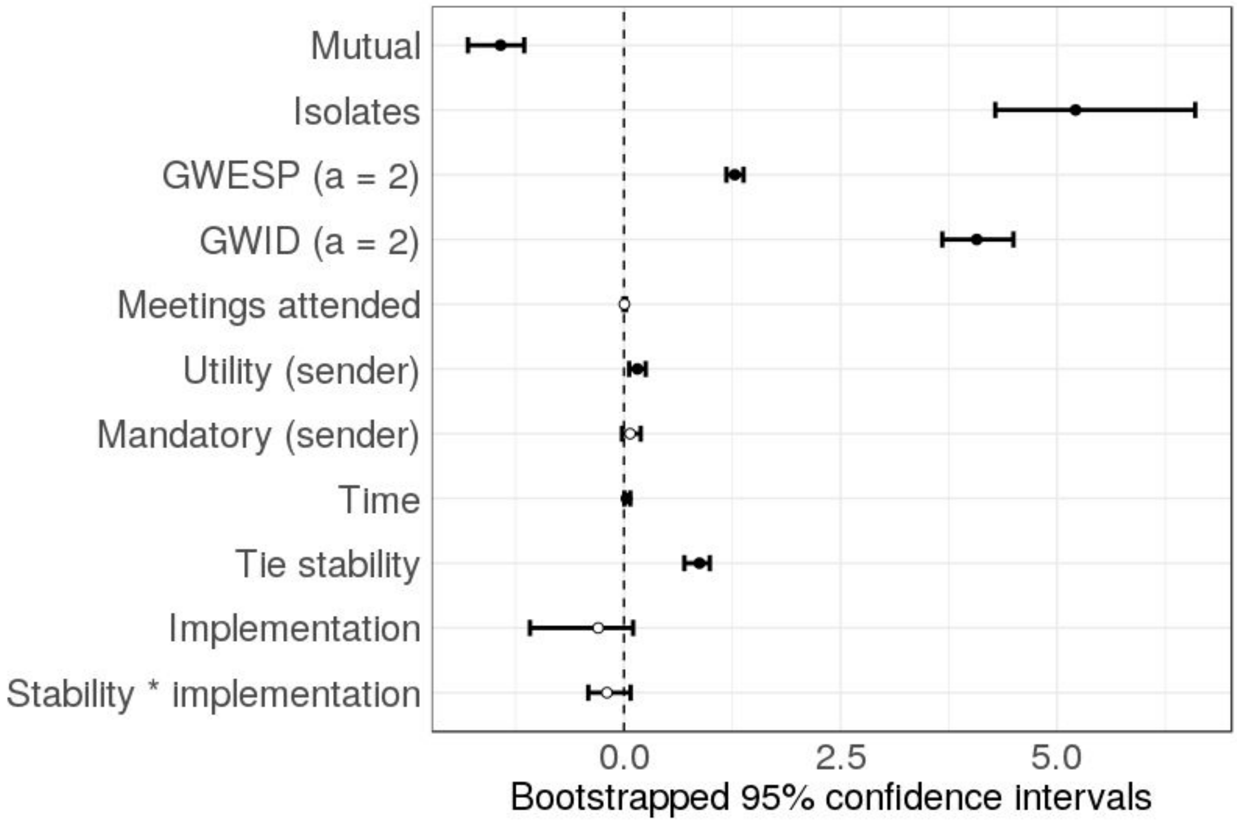

Figure 6: Model results comparing tie stability pre- and post-implementation

Finally, figure 7 presents results from a model testing the expectation of hypothesis 5, that utility actors will become less likely to take on a leadership role as the process proceeds. Actors affiliated with the utility are more likely, on average, to be a "sender" of interaction ties (these are actors who present, comment, and discuss issues at meetings). However, this effect is not shown to change as a function of time. In other words, the leadership role of the utility remains fairly consistent throughout all phases of the project.

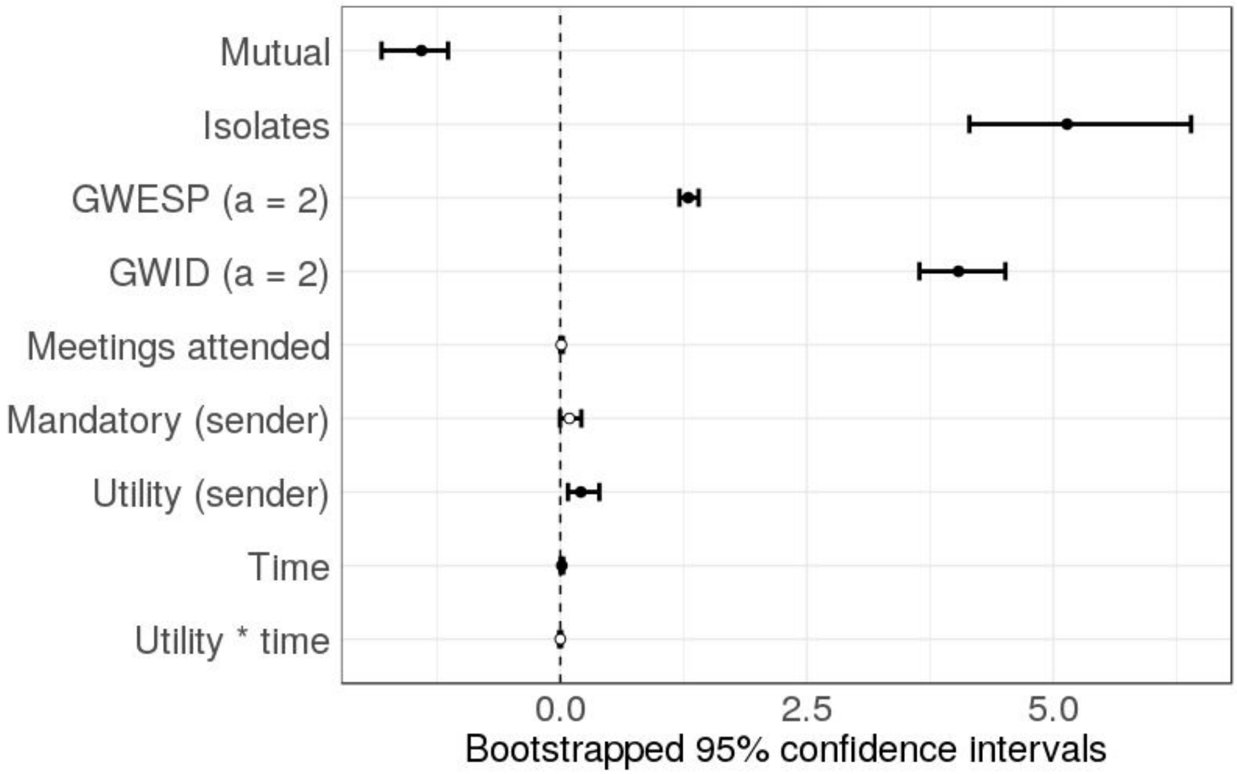

Figure 7: Model results comparing utility centrality over time

Discussion

Hypotheses 1, 2, and 5 each relate to power dynamics within the network, asking which types of organizations occupy central roles and whether central and peripheral players change their attendance and interaction patterns over time. We find support for H1, which hypothesized that the utility and mandatory conditioning agencies would occupy more central nodes throughout the relicensing. In other words, presentations and discussions are more likely to be led by the utility, the US Forest Service, the National Marine Fisheries Service, or the Washington Department of Ecology than by other attendees. This has important implications for the functioning of the collaborative process, as this subset of "leading" organizations is therefore going to drive the content of deliberation: which topics are considered during a meeting, and therefore how resources are actually managed according to the new license.

We likewise find evidence for H2, that peripheral players--those who are receiving rather than sharing information--are more likely to drop out of the relicensing over time. This supports the idea that these individuals are engaged in an ecology of games (Lubell et al. 2010), being interested enough to attend and learn about the process but not finding enough potential benefit to continue engaging. However, H2a, which posited that peripheral players from poorer or less technically-savvy organizations were especially likely to drop off, was not supported. This is an important finding, as it suggests that there was not differential access to the relicensing for community groups, tribes, and non-profits. Rather, these groups were engaging in a similar benefit-cost analysis and had the resources to continue engaging if they chose to.

Finally, H5 was not supported. This hypothesis stated that the relicensing would grow to be more inclusive over time, with the utility (and mandatory conditioning authorities) playing a less central role as other organizations began to share the lead.

Taken together, these findings paint a complex picture of the governance network in this highly collaborative relicensing. The network had distributed leadership, with both the utility and the mandatory conditioning agencies directing presentations and conversations. This makes the relicensing distinct from one where the utility alone drove most conversation, which along most common metrics would not be considered either collaborative or inclusive (Ulibarri 2015b, Quick & Feldman 2011, Arnstein 1969). However, those leadership roles were constrained to a handful of organizations relative to the full suite that participated at some point during the relicensing . And, the core network was fairly static, as other organizations did not appear to share those central nodes over time (as we would expect if the network moved toward a more equal distribution of decision-making power). Meanwhile, peripheral participants were dropping out of the network with each subsequent step of the relicensing.

That certain players are made more central because regulations say FERC has to listen to them poses an interesting governance conundrum. On one hand, the collaborative governance literature generally lauds attempts to broaden access to decision-making, as it is seen to result in fairer and potentially more durable decisions (Ansell & Gash 2008, Blair and Janousek 2013, others?). Thus, if we're trying to create a truly collaborative forum, it seems that reducing power disparities would be beneficial. This is especially true when paired with the observation that participants that are not central are more likely to leave the process. By removing the legal authority of mandatory conditioning, FERC could potentially distribute access to lead decision-making more equally across participants. At the same time, there are numerous hydropower relicensings that are not collaborative like Baker, but instead engage in the minimum required consultation. By having a handful of agencies with fairly broad interests with whom the utility must engage, FERC ensures that at a minimum, public land, anadromous fish, and water quality are adequately protected (even if that means organizations with other interests have less access to the table).

Hypotheses 3 and 4 draw on the network governance literature to predict how overall patterns of interaction would change as the network evolved. Interestingly, neither hypothesis was supported, as we did not observe a shift toward a more stable network either in the move from planning/scoping to application development nor from application review to license implementation. As one of the first quantitative longitudinal studies of a governance network, this is a critical finding for refining our theories of how networks evolve over time. Existing frameworks generally argue that grassroots networks (networks that do not stem from a policy mandate) start out fairly turbulent, with large amounts of turnover, then move to a more stable phase as relationships solidify, and finally move to a very stable, formalized (Provan & Kenis, Imperial & Koontz, Mandell & Keast). However, in this case of mandated coordination with optional collaboration, the structure of interactions did not change much over time. Where we did see change, however, was on the periphery, as participants that were not central to begin with left the network over time.

Our analysis for H4, which focused explicitly on how the network changed after there was no longer an external mandate to work together, revealed that the network became less stable after the license was issued. While a more stable network is not necessarily more effective, this does suggest that there is additional work for the convening organization to maintain the network once they no longer have external policy support. Given the high level of collaboration that Puget Sound Energy opted to create for their relicensing, we can presume that they are a fairly committed convener. If they struggled to maintain the network, a convener who is less committed to enabling collaboration could see the network dissolve completely.

Methodological Reflection

This research presents an approach combining computer-based text analysis with statistical network analysis that allows us to quantitatively observe inter-stakeholder dynamics over time. By using text scraping and natural language processing, we were able to observe attendance patterns at 591 different meetings held over a sixteen year period. This provides a highly nuanced temporal dataset, especially when compared to survey methods that perhaps capture two or three snapshots over a time series. And, by measuring a tie based on an observed interaction rather than meeting co-attendance, we gain information on dynamics within the network--who is shaping versus receiving the dialogue, and how that changes over time (Ulibarri and Scott 2017). This approach of coding ties also provides an analytical benefit, as a non-directed co-attendance model poses a computational challenge in that the observed network at a given point consists of a complete subgroup (ties between every attendee) and an empty subgroup (of non-attendees).

More generally, as described above, most public management scholarship concerning collaborative governance networks relies upon survey-based assessments (e.g., Ingold and

Leifeld 2016; Adam D. Henry et al. 2011; Shrestha 2013; Calanni et al. 2015). The subjective nature of these measures can also make it difficult to make relative comparisons between network actors. For instance, if a survey questions asks respondents to list up to five other actors with whom they collaborate, these responses are not able to reflect the relative strength or frequency of interactions. Text analysis methods thus provide an objective measure of network interactions that complement to these subjective assessments. As the first application of this text-based approach, the framework we present is fairly simple. Future research will be able to build upon this framework to test more nuanced theories. For instance, by treating presentations or other leadership actions differently than questions or comments, we could better observe who is shaping versus receiving content at the meetings and whether these meetings are more or less deliberative (Ulibarri 2015b). This could then be paired with topic modeling to see how changes in the network correspond to actual changes in the subjects being discussed. Finally, our "dynamics" primarily focus on who is engaging with whom, but there may be ways to observe other important collaborative dynamics such as trust (Emerson and Nabatchi 2015a).

Going forward, research that draws upon both objective and subjective measures of collaboration and stakeholder interactions represents an important avenue for understanding collaborative governance networks. Survey measures have important benefits that text analysis cannot fully replace. Many of the theoretical concepts understood to be key drivers of network structure and function, such as trust, reputation, and social capital (Berardo and Scholz 2010; Provan and Kenis 2008), are fundamentally subjective.

Survey instruments and other means of assessing the perspectives of network actors can garner specific information about concepts such as influence or collaborative preferences that otherwise must be inferred. Text analysis of procedural documents and other sources offers a way to ground these assessments in a framework of concrete actions and behaviors.

Conclusion

While collaborative governance networks have been studied using many different approaches within the public policy and management literatures, few studies have been able to quantitatively assess collaborative governance networks in detail over a long time period. This paper demonstrates the use of natural language processing as a way to leverage documents generated as part of a planning and management process; automated coding of these documents is a way to overcome some of the vexing data collection and measurement challenges that can hinder the study of complex governance networks (Thomas and Koontz 2011). Natural language processing is a powerful method of describing actions within the collaboration and opens upon many possibilities for future research given the vast array of textual data generated as part of the policy process.

One primary benefit of this approach is that it allows us to evaluate changes to the structure and function of a collaborative governance network over a 16-year period. Collaborative governance is normatively popular because it is considered to be way to better involve relevant stakeholders in planning and policymaking, thereby improving the quality and equity of decision-making (Scott and Thomas 2016). However, stakeholders within complex institutional environments typically have many competing demands on their time (Lubell 2013) and are unlikely to persist in their involvement if the process does not provide individual benefits of some sort (Emerson and Nabatchi 2015b). By evaluating patterns of stakeholder interaction over the course of a long term collaborative process, this paper addresses several important questions related to how participation changes over the course of a collaborative process, and how particular types of actors and the leadership roles that they plan factor into evolving network structures. We show that those mandated with convening the group remained central within the network throughout the process. Stakeholders actively contributing to the collaborative process remained relatively consistent in their participation, while less active contributors were more likely to drop out as time progressed. However, there was no evidence that less powerful groups were disproportionately impacted by these dynamics.

Taken as a whole, this research supports the assertion that leadership is critical not only for formation of collaborative partnerships but for governing the patterns of interaction that develop. The partial dissolution of the group following the end of the mandatory collaboration stage suggests that the mandate was critical in maintaining stability and actor participation.[14] Future work should build upon this study by analyzing several different processes in order to understand how system-level drivers (e.g., whether collaboration is

[14] It is also worth noting that given the statutory nature of the Baker collaborative group, the fact that the network was shown to be highly stable over time a policy output by which their success could be measured—this differs somewhat from cases of grassroots collaboration (Emerson and Nabatchi 2015).

mandate or not) affect temporal behavior. It is also worth noting that the Baker case has been documented as an instance of successful collaborative (Ulibarri 2015b); future scholarship on less successful collaborations could consider how network evolution may vary as performance of the collaborative changes.

References

Ansell, Chris, and Alison Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal of Public Administration Research and Theory* 18 (4): 543–71.

Berardo, Ramiro. 2014. "Bridging and Bonding Capital in Two-Mode Collaboration Networks." *Policy Studies Journal* 42 (2): 197–225.

Berardo, Ramiro, and John T. Scholz. 2010. "Self-Organizing Policy Networks: Risk, Partner Selection, and Cooperation in Estuaries." *American Journal of Political Science* 54 (3): 632–49.

Brin, Sergey, and Lawrence Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine - ScienceDirect." Elsevier. http://www.sciencedirect.com/science/article/pii/S016975529800110X.

Bryson, John M., Barbara C. Crosby, and Melissa Middleton Stone. 2006. "The Design and Implementation of Cross-Sector Collaborations: Propositions from the Literature." *Public Administration Review* 66 (s1): 44–55.

Calanni, John C., Saba N. Siddiki, Chris M. Weible, and William D. Leach. 2015. "Explaining Coordination in Collaborative Partnerships and Clarifying the Scope of the Belief Homophily Hypothesis." *Journal of Public Administration Research and Theory* 25 (3): 901–27.

Cameron, Kim S., and David A. Whetten. 1981. "Perceptions of Organizational Effectiveness Over Organizational Life Cycles." *Administrative Science Quarterly* 26 (4). [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University]: 525–44.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association* 19 (1): 66–86.

———. 2016. "A Critique of Dyadic Design." *International Studies Quarterly: A Publication of the International Studies Association*, March. The Oxford University Press, sqw007.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal, Complex Systems* 1695 (5). necsi.edu: 1–9.

Czarna, Anna Z., Philip Leifeld, Magdalena Śmieja, Michael Dufner, and Peter Salovey. 2016. "Do Narcissism and Emotional Intelligence Win Us Friends? Modeling Dynamics of Peer Popularity Using Inferential Network Analysis." *Personality & Social Psychology*

Bulletin, September. journals.sagepub.com. doi:10.1177/0146167216666265.
Desmarais, Bruce A., and Skyler J. Cranmer. 2012a. "Micro-Level Interpretation of Exponential Random Graph Models with Application to Estuary Networks." *Policy Studies Journal* 40 (3): 402–34.
———. 2012b. "Statistical Mechanics of Networks: Estimation and Uncertainty." *Physica A: Statistical Mechanics and Its Applications* 391 (4). Elsevier: 1865–76.
Emerson, Kirk, and Tina Nabatchi. 2015a. *Collaborative Governance Regimes*. Washington, D.C.: Georgetown University Press.
———. 2015b. "Evaluating the Productivity of Collaborative Governance Regimes: A Performance Matrix." *Public Performance & Management Review* 38 (4): 717–47.
Friedkin, Noah E., and Eugene C. Johnsen. 2014. "Two Steps to Obfuscation." *Social Networks* 39 (October). Elsevier: 12–13.
Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. 2014. *Ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (http://www.statnet.org). CRAN.R-project.org/package=ergm.
Heaney, Michael T., and Philip Leifeld. 2016. "Contributions by Interest Groups to Lobbying Coalitions." *Sites.lsa.umich.edu*. http://sites.lsa.umich.edu/mheaney/wp-content/uploads/sites/38/2016/11/Heaney_and_Leifeld-2016-11-03.pdf.
Henry, Adam D., Mark Lubell, and Mike McCoy. 2011. "Belief Systems and Social Capital as Drivers of Policy Network Structure: The Case of California Regional Planning." *Journal of Public Administration Research and Theory* 21 (3): 419–44.
Henry, Adam Douglas, and Thomas Dietz. 2016. "Policy Learning in an Evolving Environmental Risk Policy Network."
Hunter, David R. 2007. "Curved Exponential Family Models for Social Networks." *Social Networks* 29 (2): 216–30.
Hunter, David R., and Mark S. Handcock. 2006. "Inference in Curved Exponential Family Models for Networks." *Journal of Computational and Graphical Statistics* 15 (3): 565–83.
Imperial, Mark T., Erik Johnston, Melinda Pruett-Jones, Kirsten Leong, and Jennifer Thomsen. 2016. "Sustaining the Useful Life of Network Governance: Life Cycles and Developmental Challenges." *Frontiers in Ecology and the Environment* 14 (3): 135–44.
Imperial, Mark T., and Tomas Koontz. 2007. "Evolution of Collaborative Organizations For Watershed Governance: Structural Properties, Life-Cycles, and." people.uncw.edu. http://people.uncw.edu/imperialm/Instructor/Papers/APPAM_07_Imperial_Koontz_Final_11_5_07.pdf.
Ingold, Karin, and Philip Leifeld. 2016. "Structural and Institutional Determinants of Influence Reputation: A Comparison of Collaborative and Adversarial Policy Networks in Decision Making and Implementation." *Journal of Public Administration Research and Theory* 26 (1): 1–18.
Jasny, Lorien. 2012. "Baseline Models for Two-Mode Social Network Data." *Policy Studies Journal* 40 (3): 458–91.
Koskinen, Johan, and G. Daraganova. 2013. "Exponential Random Graph Model Fundamentals." In *Exponential Random Graph Models for Social Networks: Theory,*

Methods, and Applications, edited by Dean Lusher, Johan Koskinen, and Garry Robins, 49–76. New York, NY: Cambridge University Press.
Koskinen, Johan, and Tab Snijders. 2013. "Simulation, Estimation, and Goodness of Fit." *Exponential Random Graph Models for Social Networks*. Cambridge University Press Cambridge, 141–66.
Leifeld, Philip, Skyler J. Cranmer, and Bruce A. Desmarais. 2015. "Estimating Temporal Exponential Random Graph Models by Bootstrapped Pseudolikelihood." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.432.4592&rep=rep1&type=pdf.
Leifeld, Philip, Skyler J. Cranmer, Bruce A. Desmarais, and Maintainer Philip Leifeld. 2015. "Package 'xergm.'" 143.107.212.50. http://143.107.212.50/web/packages/xergm/xergm.pdf.
Levy, Michael A. 2016. "Gwdegree: Improving Interpretation of Geometrically-Weighted Degree Estimates in Exponential Random Graph Models." *The Journal of Open Source Software* 1 (3). doi:10.21105/joss.00036.
Levy, Michael A., Mark Lubell, Philip Leifeld, and Skyler Cranmer. 2016. "Estimating and Interpreting Geometrically Weighted Statistics in Exponential Random Graph Models." In .
Lubell, Mark. 2013. "Governing Institutional Complexity: The Ecology of Games Framework." *Policy Studies Journal* 41 (3): 537–59.
Lubell, Mark, Adam D. Henry, and Mike McCoy. 2010. "Collaborative Institutions in an Ecology of Games." *American Journal of Political Science* 54 (2): 287–300.
Lubell, Mark, Garry Robins, and Peng Wang. 2014. "Network Structure and Institutional Complexity in an Ecology of Water Management Games." *Ecology and Society* 19 (4): 23.
Lubell, Mark, John T. Scholz, Ramiro Berardo, and Garry Robins. 2012. "Testing Policy Theory with Statistical Models of Networks." *Policy Studies Journal* 40 (3): 351–74.
Lusher, Dean, Johan Koskinen, and Garry Robins. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. New York, NY: Cambridge University Press.
Mandell, Myrna P., and Robyn Keast. 2008. "Evaluating the Effectiveness of Interorganizational Relations through Networks: Developing a Framework for Revised Performance Measures." *Public Management Review* 10 (6): 715–31.
Milward, H. Brinton, and Keith G. Provan. 2000. "Governing the Hollow State." *Journal of Public Administration Research and Theory* 10 (2). PMRA: 359–80.
Park, Hyun Hee, and R. Karl Rethemeyer. 2014. "The Politics of Connections: Assessing the Determinants of Social Structure in Policy Networks." *Journal of Public Administration Research and Theory* 24 (2). Oxford University Press: 349–79.
Provan, Keith G., and Patrick Kenis. 2008. "Modes of Network Governance: Structure, Management, and Effectiveness." *Journal of Public Administration Research and Theory* 18 (2): 229–52.
Purdy, Jill M. 2012. "A Framework for Assessing Power in Collaborative Governance Processes." *Public Administration Review* 72 (3). Blackwell Publishing Ltd: 409–17.
R Core Team. 2016. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robins, Garry, Jenny M. Lewis, and Peng Wang. 2012. "Statistical Network Analysis for Analyzing Policy Networks." *Policy Studies Journal* 40 (3): 375–401.

Sabatier, Paul A., William D. Leach, Mark Lubell, and Neil W. Pelkey. 2005. "Theoretical Frameworks Explaining Partnership Success." In *Swimming Upstream: Collaborative Approaches to Watershed Management*, edited by Paul A. Sabatier, W. Focht, Mark Lubell, Z. Trachtenberg, A. Vedlitz, and M. Matlock, 173–200. Cambridge, MA: MIT Press.

Schuler, Karin Kipper. 2005. "VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon." University of Pennsylvania. http://repository.upenn.edu/dissertations/AAI3179808/.

Scott Long, J., and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata, Second Edition*. Stata Press.

Scott, Tyler A. 2016. "Analyzing Policy Networks Using Valued Exponential Random Graph Models: Do Government-Sponsored Collaborative Groups Enhance Organizational Networks?" *Policy Studies Journal* 44 (2): 215–44.

Scott, Tyler A., and Craig Thomas. 2016. "Unpacking the Collaborative Toolbox: Why and When Do Public Managers Choose Collaborative Governance Strategies?" *Policy Studies Journal*. doi:10.1111/psj.12162.

Shrestha, Manoj K. 2013. "Self-Organizing Network Capital and the Success of Collaborative Public Programs." *Journal of Public Administration Research and Theory* 23 (2): 307–29.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36 (1): 99–153.

Thomas, Craig W., and Tomas M. Koontz. 2011. "Research Designs for Evaluating the Impact of Community-Based Management on Natural Resource Conservation." *Journal of Natural Resources Policy Research* 3 (2): 97–111.

Ulibarri, Nicola. 2015a. "Collaboration in Federal Hydropower Licensing: Impacts on Process, Outputs, and Outcomes." *Public Performance & Management Review* 38 (4): 578–606.

———. 2015b. "Tracing Process to Performance of Collaborative Governance: A Comparative Case Study of Federal Hydropower Licensing." *Policy Studies Journal* 43 (2): 283–308.

Ulibarri, Nicola, and Tyler A. Scott. 2017. "Linking Network Structure to Collaborative Governance." *Journal of Public Administration Research and Theory* 27 (1): 163–81.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Vol. 8. New York, NY: Cambridge University Press.

Weber, Edward P. 2009. "Explaining Institutional Change in Tough Cases of Collaboration:'ideas' in the Blackfoot Watershed." *Public Administration Review* 69 (2): 314–27.

Appendix A: Comparison of automated text analysis and hand coded ties

In order to better understand the nature of the data generated by our automated text analysis, we hand-coded a sample of 49 randomly selected meetings for comparison. Figure A1 orders these 49 meetings along the X-axis by the number of attendees observed via hand-coding. For each of these meetings, we then plot the number of attendees observed via hand-coding (hollow, darkly bordered points) and the number of attendees observed via automated coding (solid, lightly colored points).

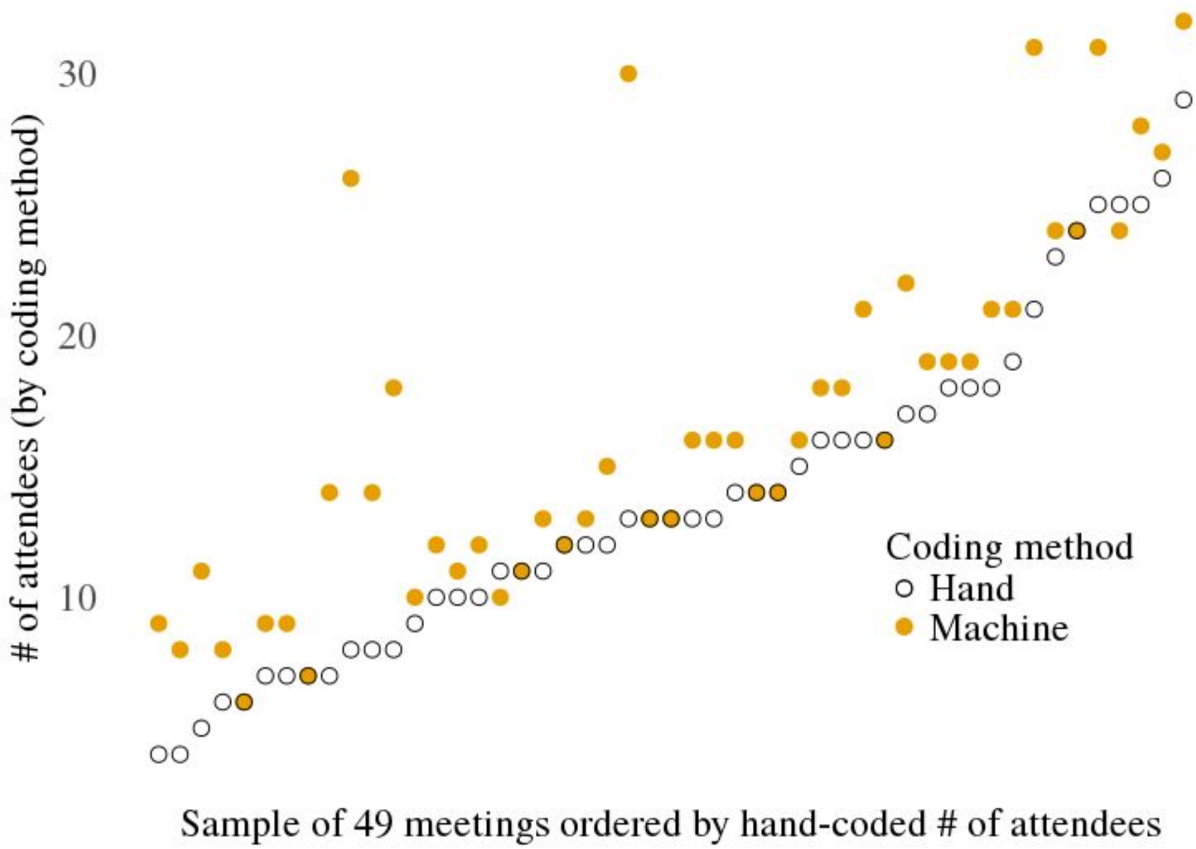

Figure A1: Comparison of hand-coded and machine-coded meeting attendance

Figure A1 shows that for the most part, automated attendance tracks fairly closely to hand-coded attendance. With two exceptions where the machine-coded number is one unit lower than the hand-coded number, the machine-coded data appear to be most subject to false positives. This is to be expected, since text processing algorithms are excellent at extracted named entities found in text (and thus in general not prone to overlook a name and generate a false negative), but contextualizing the reason why said name occurs in the text is more challenging. An examination of the cases shown in figure A1 where the machine-coded attendance number far outpaces the hand-coded number revealed that these meeting summary documents contained a both a list of those present at the meeting and a list of individuals on a "distribution list" who did not attend the meeting but desired to be kept apprised of the process. On a per-meeting basis, the average difference between hand-coded attendance and machine-coded attendance is 2.96 more machine coded attendees. As the average number of attendees per meeting within the 49 meeting subsample is 13.8, machine-coding on average over-inflates attendance by around 21%.

It is also important to note that the machine-coded attendance data are not purely the raw extracted named entities. After performing entity extraction (as described in the methods section above), we took a series of basic data cleaning steps (the code for which are available as part of the data and code for this analysis made publicly available). These steps included filtering out obvious false positives that did not correspond to actual names. For example, terms such as “Dolly Varden” (a trout species native to the region) and “Howard Hansen” (the name of another dam in the region) are incorrectly recognized as named persons. Further, the structure of the meeting summary documents poses a challenge in many cases, for instance, where participation is itemized such that an entry might say something akin to: “Tony Revise document and provide new draft next month.” Given the capitalization of “Revise”, this is prone to being extracted as an entity named “Tony Revise”. Thus, we also filter out extracted entities that contain actions such as “revised”, “draft”, “email”, “clean”, and “report” (after carefully checking to ensure that any words used as filters do not also happen to be the name of a participating stakeholder). We also corrected for misspelled names with code script that used regular expression searches to identify pairs of extracted named entities that differed by a limited number of character changes, additions, or removals.

While attendance data are used to identify tie recipients in the model, participation data are needed to code where ties originate. We conduct a similar comparison exercise using the hand-coded meeting sample. Figure A2 compares the number of hand-coded participation actions observed for each meeting with the number of participation actions observed via our automated text analysis. For the most part, we again see that machine-coded observations do track reasonably well with hand-coded observations. However, whereas automated coding of attendance primarily generated false positives, figure A2 shows that automated coding of participation appears more prone to producing false negatives (i.e., failing to recognize a stakeholder’s participation action). The most prominent driver of this is that that hand-coding is better able to disambiguate names and to identify the subject(s) of pronouns. In other words, a contextualized reading of a document increases the chance that a recognized participation action can be clearly attributed to a specific attendee. Thus, the machine-coding approach exhibits a downward bias in terms of observing participation.

On a per-meeting basis, the average difference between the number of hand-coded participation actions and machine-coded participation actions is 4.3 more hand-coded observations. The average number of total participation actions observed per meeting in the hand-counted sample is 22.73; machine-coding thus captures about 81% of total participation per meeting on average.

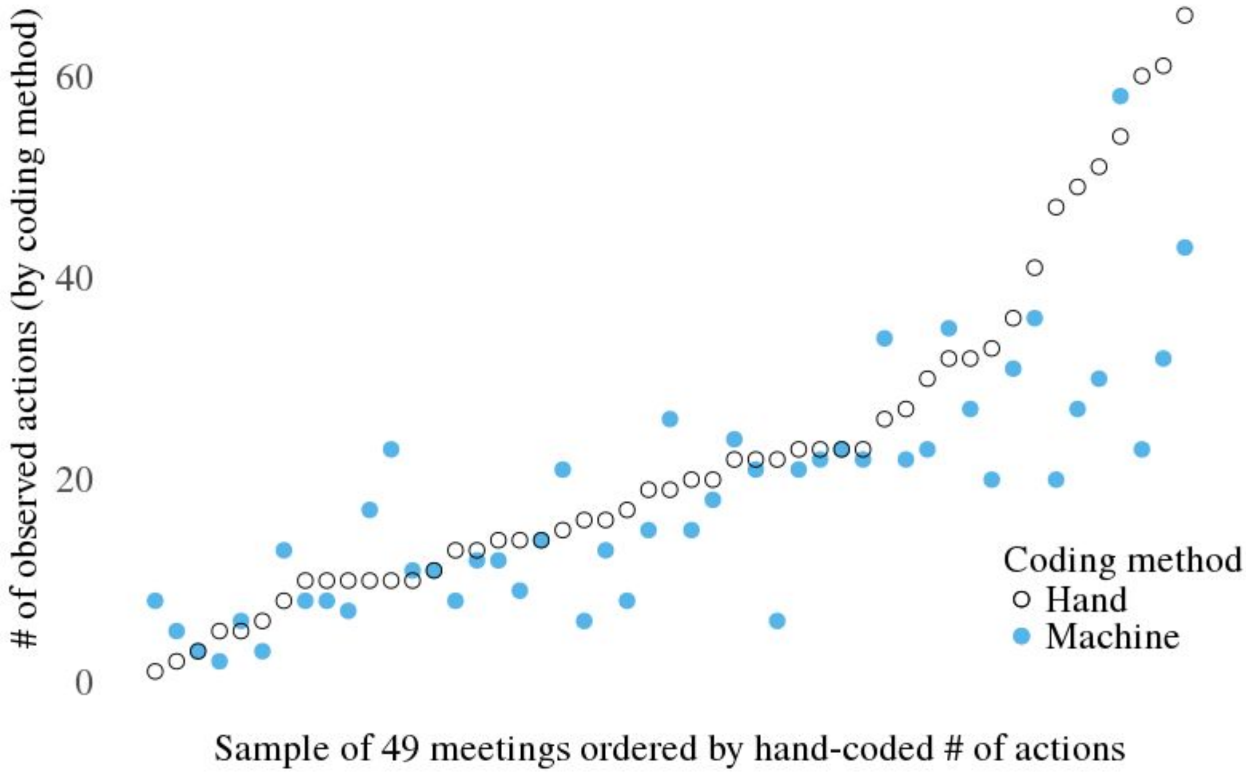

Figure A2: Comparison of hand-coded and machine-coded meeting participation actions

While figure A2 presents the total count of all participation actions observed for each meeting via hand-coding and automated-coding methods, recall that network ties are coding in binary fashion for this analysis. In other words, if an attendee is documented in a meeting summary to have given a presentation, provided a comment, and made a suggestion (i.e., three separate participation actions), they are only assigned one "interaction" tie stemming to other attendees. Thus, with respect to understanding how the coding method used might influence analysis results, it is important to consider how the hand-coded and automatically coded data differ with respect to the number of unique stakeholders observed to participate in a given meeting. In other words, for the current analysis is it does not really matter if the hand-coded and automatically coded participation data differ in terms of how many times an attendee participation; rather, what can influence the results is whether an attendee is recorded as having participated at all.

In terms of capturing unique individual meeting participants, the automated coding strategy performs very well. Figure A3 compares hand-coded and automatically coded unique participants by meeting; while it might appear that there is considerable divergence, note that the y-axis is shrunken considerably relative to figure A2. No individual meeting differs by more than 5 observations between the two codings, and the average difference between hand-coded and machine-coded observations is 0.38 observations per meeting. The average number of actions taken by unique stakeholders per meeting is 6.35, meaning that machine coding captures about 94% of unique participation on average.

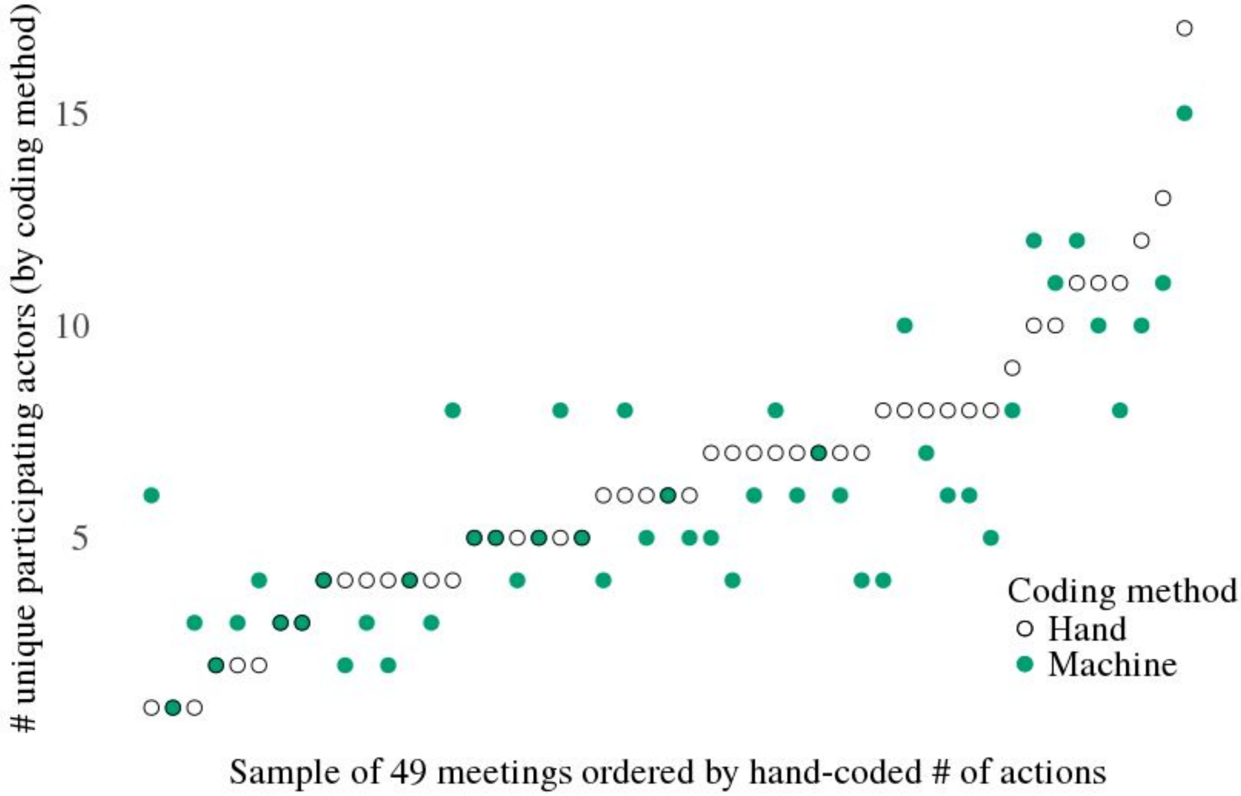

Figure A2: Comparison of hand-coded and machine-coded participating meeting actors

Appendix B: Supplementary model results

Table B1: Tabular results for models used to test H1, H2, and H2A

	Model 1	Model 2	Model 2A
Edges	-11.046*	-11.203*	-11.126*
	[-11.804; -10.229]	[-12.066; -10.370]	[-11.943; -10.194]
Mutual	-1.409*	-1.448*	-1.446*
	[-1.772; -1.135]	[-1.846; -1.165]	[-1.810; -1.174]
Isolates	5.140*	5.409*	5.411*
	[4.197; 6.472]	[4.441; 6.560]	[4.516; 6.613]
Meetings attended	0.010	0.036*	0.036*
	[-0.001; 0.025]	[0.024; 0.057]	[0.023; 0.055]
GWID (a = 2)	4.039*	4.665*	4.664*
	[3.616; 4.563]	[4.266; 5.144]	[4.271; 5.109]
GWESP (a = 2)	1.298*	1.322*	1.322*
	[1.207; 1.395]	[1.226; 1.427]	[1.218; 1.423]
Mandatory (sender)	0.092		
	[-0.001; 0.199]		
Utility (sender)	0.204*		
	[0.109; 0.298]		
Time	0.015*	0.025*	0.025*
	[0.004; 0.027]	[0.015; 0.039]	[0.014; 0.037]
High resource [HR]		0.172*	0.047
		[0.088; 0.257]	[-0.160; 0.236]
Prior centrality		-2.303*	-2.661*
		[-2.825; -1.927]	[-3.406; -2.129]
HR * Prior centrality			0.522
			[-0.123; 1.305]

Table B2: Tabular results for models used to test H3, H4, and H5

	Model 3	**Model 4**	**Model 5**
Edges	-10.498*	-11.057*	-11.047*
	[-11.310; -9.663]	[-11.938; -10.187]	[-11.874; -10.231]
Mutual	-1.426*	-1.425*	-1.409*
	[-1.778; -1.142]	[-1.802; -1.152]	[-1.813; -1.139]
Isolates	5.222*	5.215*	5.140*
	[4.315; 6.488]	[4.288; 6.596]	[4.149; 6.397]
Meetings attended	0.006	0.005	0.010
	[-0.004; 0.024]	[-0.005; 0.022]	[-0.001; 0.026]
GWID (a = 2)	4.070*	4.073*	4.038*
	[3.617; 4.519]	[3.676; 4.497]	[3.642; 4.514]
GWESP (a = 2)	1.278*	1.279*	1.298*
	[1.182; 1.379]	[1.183; 1.380]	[1.208; 1.402]
Mandatory (sender)	0.070	0.070	0.092
	[-0.032; 0.201]	[-0.023; 0.193]	[-0.004; 0.213]
Utility (sender)	0.149*	0.154*	0.208*
	[0.047; 0.250]	[0.059; 0.250]	[0.078; 0.395]
Time	-0.001	0.028*	0.015*
	[-0.047; 0.021]	[0.003; 0.073]	[0.003; 0.029]
Post-planning	0.222		
	[-0.080; 1.066]		
Dyad stability	0.457*		
	[0.256; 0.531]		
Stability * post-planning	-0.094		
	[-0.196; 0.112]		
Implementation		-0.298	
		[-1.086; 0.105]	
Tie stability		0.871*	
		[0.696; 0.990]	
Stability * implementation		-0.197	
		[-0.413; 0.076]	
Utility * time			-0.000
			[-0.012; 0.009]

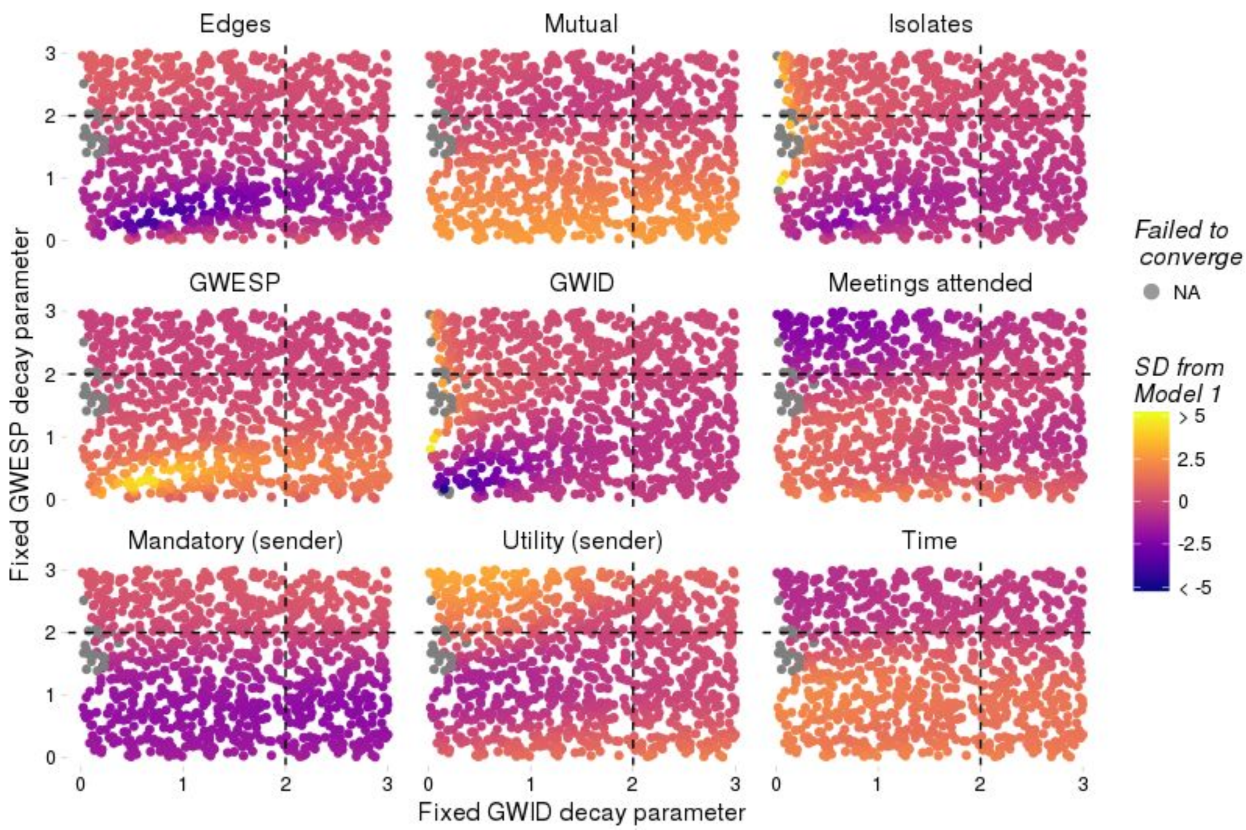

Figure B1: Standardized coefficient estimates for 1000 simulations varying fixed decay parameters for GWID and GWESP terms.

As discussed in the main body of the paper, the choice of decay parameters used to fit the GWID and GWESP terms can influence resultant model estimates (Levy et al. 2016). In order to gauge how fixing each decay parameter at $\alpha_{GWESP} = \alpha_{GWID} = 2$ might unduly influence our results, we conducted 1000 iterations of randomly selecting decay parameters between 0.01 and 3 (as noted by Levy et al. (2016), a 0 value is undesirable) for the GWID and GWESP terms (independently for each term) and refitting the model. We then calculate the difference between each parameter estimate and the estimate shown in table B1 for model 1, and then divide this difference by the standard deviation of the 1000 estimations generated for each parameter. Figure B1 plots the distribution of parameter estimates by GWID and GWESP shape parameter, where the shading for an estimate at a given α_{GWESP} and α_{GWID} value reflects the difference, in standardized units, between that the coefficient estimate at that decay parameter combination and the estimated coefficient at $\alpha_{GWESP} = \alpha_{GWID} = 2$. As shown in figure B1, estimates are generally consistent at values close to the $\alpha_{GWESP} = 2$ and $\alpha_{GWID} = 2$ values used for our primary results.

On additional way to assess the impact of α_{GWESP} and α_{GWID} is to examine how the sign and statistical significance of each parameter changes as these decay parameters are permuted. Figure B2 shows that the choice of α_{GWESP} does play a role in the interpretation of estimated coefficients. There is a stark change in overall model results at approximately $\alpha_{GWESP} = 1.5$, such that the coefficients for the mutual and isolates statistics, as well as the the sender effect of mandatory signatory actors, change in sign. More importantly, however, the two terms that figure prominently in testing subsequent hypotheses (the sender effect for utility affiliated actors and time effect) exhibit a consistent sign across all permutations of α_{GWESP} and α_{GWID}, indicating that the choice of these decay parameters is not unduly driving these results.

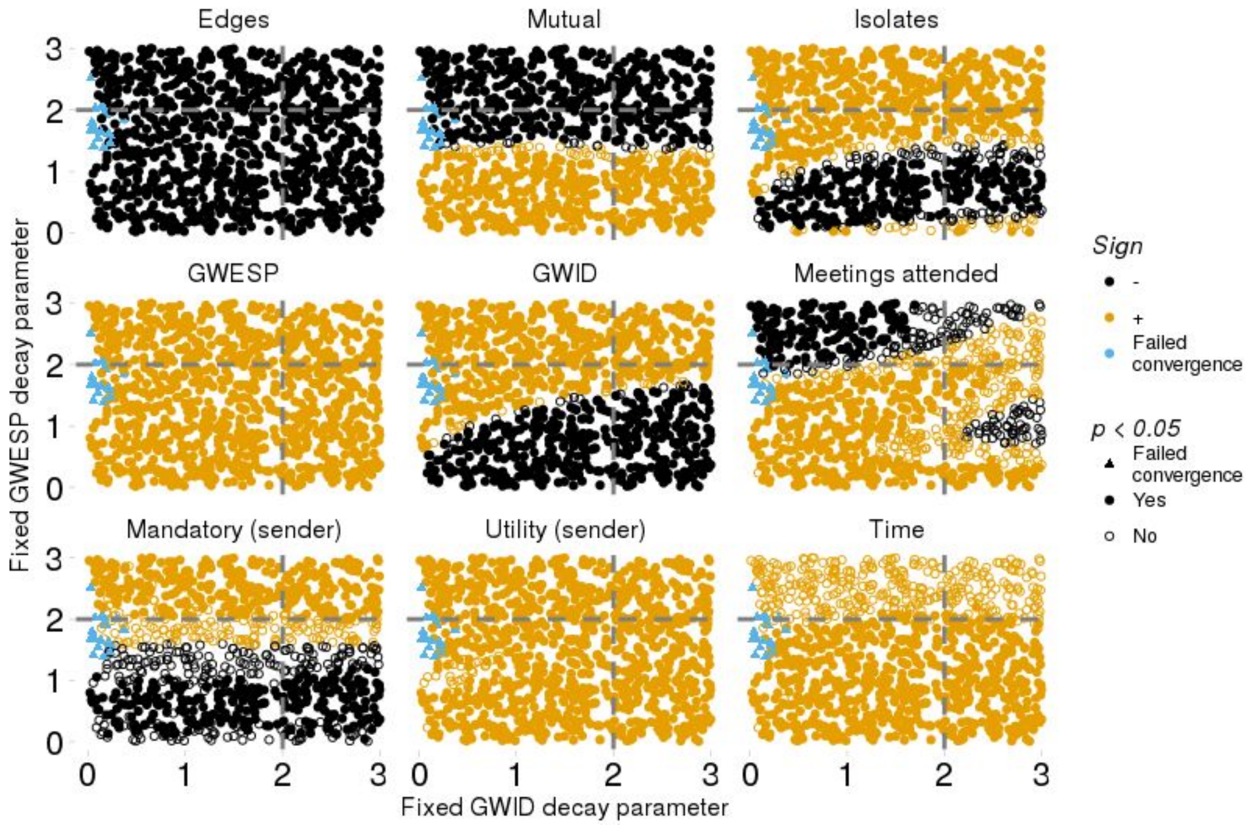

Figure B2: Sign and significance of model 1 parameters for 1000 simulations varying fixed decay parameters for GWID and GWESP terms.

Appendix C: Model goodness of fit

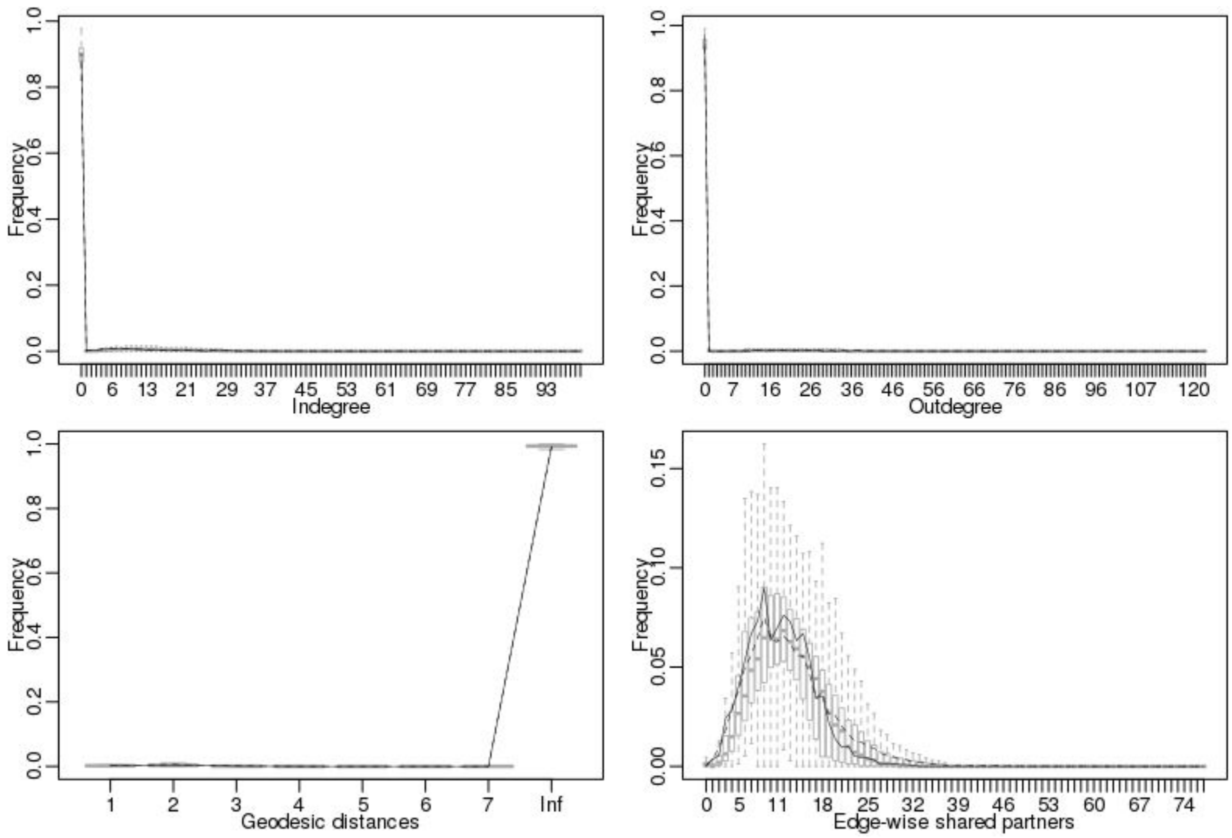

Figure C1: Comparison of observed and simulated network statistics generated by model 1.

As described in the methodology section of the manuscript, a TERGM facilitates inference by comparing the observed network to a distribution of simulated networks. Thus, it is important to confirm that the model is able to generate simulated networks that bear structural resemblance to the observed network (Koskinen and Snijders 2013). Figure C1 below compares four observed network statistics (outdegree, indegree, minimum geodesic distance [the minimum number of ties linking each pair of network actors], and edgewise shared partners) to a distribution of these same statistics based upon simulated networks generated by model 1. Overall, there is little discrepancy between the observed network (averaged across time steps) and simulated graphs. Because the overall network is quite sparse, and there are many high degree actors, the top two panels in figure A4 likely appear somewhat strange. The basic issue is that in both the simulated and observed networks, there are very few actors who have an indegree values of 45 or an outdegree value of 60. Because both the observed networks (dark line) and simulated networks have very few such actors at any given degree value, the frequency line remains at or near zero. The more important implication, however, is that the simulated networks do not have unwanted clusters of high degree actors, and that the distribution of indegree and outdegree values by node track resemble observed values.

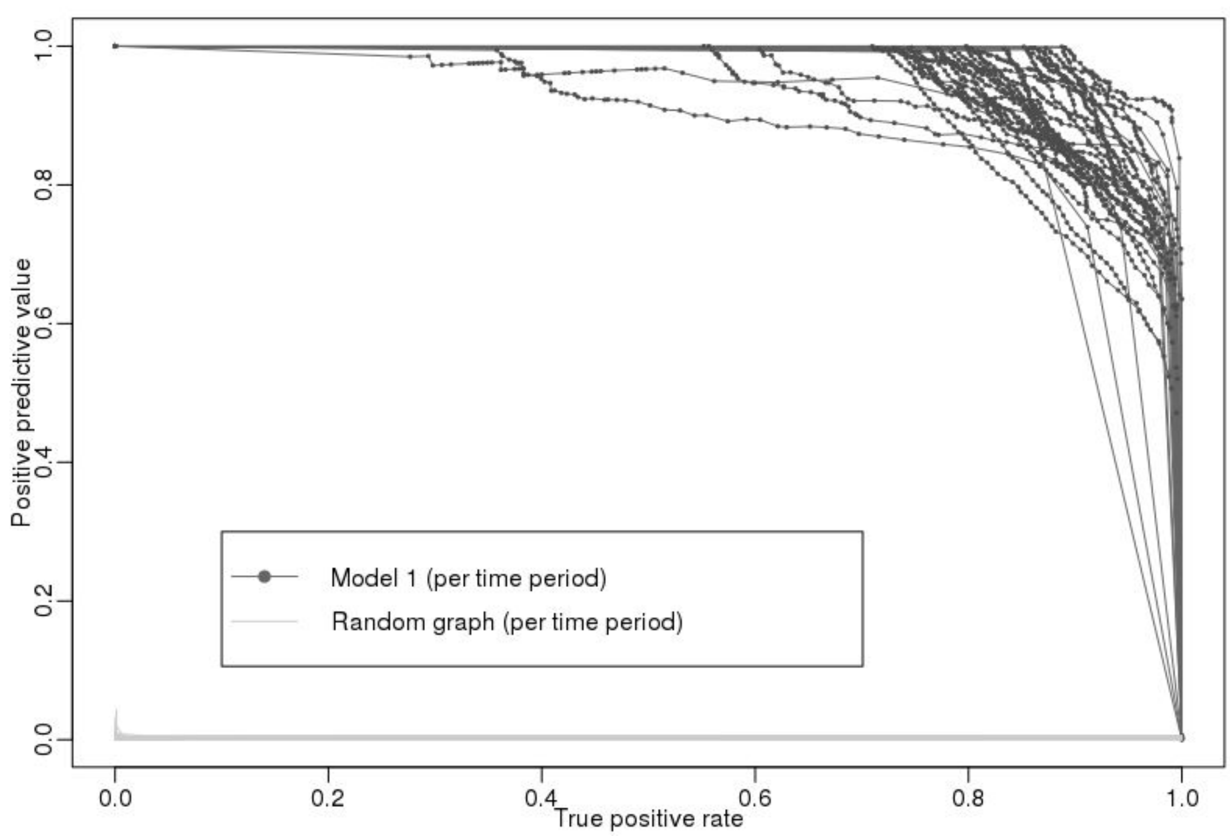

Figure C2: Precision-recall curves for model 1 and random graph by time period

A second way to gauge goodness-of-fit is to consider the predictive ability of the estimate model. Precision-recall curves compare the ability to correctly predict the presence of ties that are observed (the true positive rate is the proportion of observed ties correctly predicted by the model, i.e., "recall") against the proportion of ties predicted by the model that are actually observed in the network (the number of true positives divided by the total number of predicted ties, i.e., "precision"). As a general rule, there is a tradeoff between precision (which is increased by making the model more discerning and reducing false negatives) and recall (which is increased by making the model less discerning and reducing false negatives), such that a precision-recall curve trends downwards. Figure C2 shows that model 1 performs very well in both regards, but still exhibits this tradeoff. Most importantly, compared to the precision-recall curves plotted for random graph estimates (i.e., an intercept-only model with just an edges to to control for density), model 1 performs far better.