
Emergence of Hierarchy via Local Adaptation to Diffusion

Max Bromberg
5266570



Masterarbeit

Fachbereich Physik
der
Freie Universität Berlin

Supervisor: Dr. Philipp Lorenz-Spreen

Examiners: Prof. Dr. Roland Netz & Prof. Dr. Ana-Nicoleta Bondar

December 22nd 2020

Abstract

Though evidence of diverse community structures abound in network science and daily life, the mechanics of their formation remains obscured by their apparent complexity. In a largely computational investigation, this thesis propounds a possible theory behind the emergence of different community structures based on the behavior of a novel model in which agents make local, conserved adjustments towards a subset of their community. The model results suggests the natural emergence of hierarchy as predicated on diffusive network communication and the consistency of the community leaders, while the resulting hierarchical network specialization consistently comes at the expense of all other intermediary communication. This research was conducted with the impact of increasingly digital communications in mind, allowing for a parameter space which ideally encompasses the complete range of marginal connection costs and network connectivity, and as such may remain relevant in both its specific applications and interpolations between different communication regimes.

After a general introduction to the theoretical motivation and network science background, we consider the compiled derivation of the diffusive communication method, the Random Walker Effective Distance before detailing the construction and operation of the model. There then follows a description of the relevant network observables used to discern its structure, behavior, and corresponding efficiency, before discussing the investigated model configurations, their real-world reflections, and the resultant model behavior.

Contents

1	Introduction	1
2	Theoretical Background	3
2.1	Network Theory	3
2.2	Random Walkers	5
2.3	Construction of Random Walker Effective Distance	9
3	Model Description	16
3.1	General Model Form	16
3.2	Network Initialization	17
3.3	Model Synopsis	18
3.4	Information Seeding	19
3.5	Information Dispersal	20
3.6	Network Adaptation	21
3.7	Model Directionality	24
4	Network Observables	25
4.1	Observables	25
4.2	Classifying Network Structure	28
4.3	Global vs Diffusive Communication Efficiency	34
5	Results	38
5.1	Model Configurations	38
5.2	Model Behavior	48
6	Conclusion	61
7	Appendix	63
7.1	List of Variables	63
7.2	Data and Code Availability	63
7.3	Computational Requirements	63
7.4	Model Design Notes	64
7.5	Observable Design Notes	65

Introduction

Twice doubled in the last century, humanity's own exponential growth [1] is dwarfed by that of its associated information. Responsibility for this combinatoric explosion of man-made information lies with the digitization, and subsequent efficiency and feasibility of both recording and globally communicating the increasingly diverse and specialized set of human activities. Were these activities simply recorded and diversified apace with Moore's Law, information would likewise be constrained to compositely compounded exponential growth, however the simultaneous gains in communication efficiency provided for the natural growth of social and economic networks whose permutations of possible paths promulgated a combinatoric count of connections. Equipped with such a plethora of data and novel computational methods and resources, many predictions previously the purview of analytic answers are now readily available without recourse to theory; in the words of Chris Anderson, "*...Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.*" [2].

The notion that theory is superfluous in the large-scale limit of data and computational power exaggerates the utility of such a silicon oracle, while discounting the sacrifices it should require. Accurate, big-data predictions are of limited use without theory to guide our understanding and reactions thereto, as without ourselves possessing the requisite computational power to comprehend these coupled complex systems we must distill their behavior into their underlying, analytically generalizable mechanisms. There are practical, as well as conceptual perils to this profusion of data; "*What counts as a meaningful finding when the number of data points is so large that something will always be significant?*" [3]. How should the significance of minuscule coupled effects be aggregated or analyzed? How to prioritize pertinent problems when data availability encourages only answering and analyzing volunteered corporate data? [4] While many of these considerations have only more recently gained relevance as computational tools develop to rival the predictive power of many existing theories, attempts to analyze the root, highly coupled, systems producing these data have arguably dated to classical antiquity [5], though the more recent statistical physics and network science based endeavors date from the last 50 years. [6] Having garnered an assortment of monikers of various specificity (econophysics, sociophysics, applied complexity, dynamical systems theory, &c...) [7], these pursuits, broadly termed complexity studies, have sought to distill the highly complex, multivariate behavior of biological, economic and sociological systems into simple models which often exceed computational limitations and form analytic expressions and associated theory. The common mechanics of network or agent based models are themselves developed in more recent advances in graph theory, network science, game theory and statistical physics, while the theory behind the models themselves seek to explain the extant corpus of economic, anthropological or else technological theory and ascribe an explicit analytic interpretation thereto. Theories originating outside the natural sciences have not the luxury of perfectly replicable physical experimentation on the constants of the universe, and thus are accordingly complex, if only implicitly when viewed from the perspective of

the natural scientist who sees in notion of *group think* [8] or cliques an impossibly large set of coupled variables and confounding agent actions. It is a truth universally acknowledged, that a scientist, in possession of good data, must be in want of a theory. Only now equipped with detailed data whose predictive power confounds its analytic interpretation, do the natural sciences make their foray into explaining anthropologically synthesized phenomena.

When developing mathematical and computational models intending to reflect theory, it is paramount the complexity of the model remains minimal, or else the number of possible interpretations of the model disallows validation of the associated theory. Furthermore, if a theory precedes its model, it is possible for the modeler to, through the introduction of greater complexity, tailor the model to fit the expectations of the theory, potentially invalidating the model's relevance; or as Jonny von Neumann once put it, "...with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." [9, 10]. As increasingly digitized social networks have grown in prominence and performance, there have been considerable efforts spent in grokking, via existing network theory, their dimensions [11], behavior [12, 13, 14, 15, 16, 17, 18, 19], and implications. [20, 21, 22, 23, 24] and though there are several early explanations of similarly hierarchical¹ network formation, such as preferential attachment [26] and the Ravasz Algorithm [27], only more recently have considerations of the efficiency of these emergent network structures as a whole been considered. [28, 29, 30, 31] Much of the complexity implicit in network studies arises from the combinatoric number of paths possible for information to be routed through a network², obscuring the methods and motivations of the agents in their communication.

Networks are nonetheless formed through the aggregate actions of individual communication patterns which we aspire to grasp both locally and globally. The essential aim, as in this thesis, is to derive a basic set of governing principles for individual interactions which may reliably reproduce myriad forms of social network structure via variance in a single, readily interpretable parameter. Though unfortunately this platonically ideal model alludes the author, the model introduced nonetheless produces a wide range of network topologies and exhibits ready reaction to the governing, readily interpretable parameters.

Leaning heavily on epidemiological research concerned with the spread of viral pathogens [32, 33, 34, 35], this thesis endeavors to craft a model of structural network evolution limited via local adaptation to viral information dispersal, and analyze the efficiency of the resultant networks evolved from different parameter regimes, where the model parameters correspond to psychological heuristics. In this pursuit, we first introduce the necessary network theory [25, 36] and development of effective distance metrics [32, 34, 37, 38], followed by a description, theoretical justification and analysis of the resulting Source to Structure model, before considering its results and placing them in the context of the recently developed efficiency [31] and topological hierarchy morphospaces [29, 31, 30].

¹Many social (e.g. actor) networks are fairly hierarchical, exhibiting scale free degree distributions [25]

²The reason information routing through a network is *implicit* is that, apart from the rarity of possessing both the computational power and data fidelity to know and analyze messages (let alone their weights/impacts), most networks are reduced to their adjacency matrix in analysis, which do not contain information on individual messages.

Chapter 2

Theoretical Background

2.1 Network Theory

To best analytically contextualize networks and their relevance to this thesis' results, we here give a brief formal introduction to networks as conventionally defined via graph theory¹, beginning with static topologies. Graphs are minimal representations of coupled data, used to represent everything from webpages [25, 36] and twitter posts [12] to etymological roots [36] and protein interactions [30], and as such are fundamental tools in understanding the interaction between systems whose independent behavior is often taken a priori. At its most basic form, a graph is a pair, $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ consisting of a set of nodes \mathcal{N} (also *vertices* or *points*) and edges \mathcal{E} (also *lines* or *links*), which represent relations between the nodes. Every edge is thus defined as a tuple $(i, j) \in \mathcal{E} \subseteq \mathcal{N}^2$ joining the nodes i and j , whereby i and j are thusly considered adjacent. Any edge for which $i = j$ is considered a *self loop*. Continuous paths consisting of sequential adjacent edges leading from a source node s to a target node t ,

$$\omega(s \rightarrow t) = [(u_0, v_0), (u_1, v_1), \dots, (u_k, v_k)] \quad (2.1)$$

define a *st-walk*, where $s \equiv u_0$, $t \equiv v_k$ are the initial and final nodes, respectively.

If $\forall n_i \in \mathcal{N} \exists \omega(s \rightarrow t) \text{ s.t. } s = n_i, t = n_j \forall j \in \mathcal{N}$ then the graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ is said to be *connected*, as there is a walk from every node to every other node in the network through the existing edges, and the walk is considered an *absorbing walk* if the target is visited only once. If every edge is only crossed once through the entirety of the walk, the edge sequence is considered a *path* π , and the length of the longest shortest path between any two given nodes is considered the *network diameter*. If the network is connected and its diameter is one, i.e. all nodes are mutually adjacent, the network is said to be *complete*. As tuples are ordered pairs, the inherent order of an edge may be considered to induce directionality into the graph structure, so that an edge (i, j) does not imply the existence of its inverse, (j, i) . A most practical representation of a graph may be prescribed via an *Adjacency Matrix* \mathbf{A} :

$$(\mathbf{A})_{i,j} = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

which is therefore symmetric if $\forall (i, j) \in \mathcal{E} \exists (j, i) \in \mathcal{E}$. Naturally, graphs are not constrained to directional symmetry, and when this condition is not met (i.e. \mathbf{A} is not symmetric) the graph is said to be directed. Note that for notational simplicity, we write the number of nodes $|\mathcal{N}|$ simply as N . In this case, the

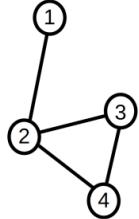
¹As graph theory and network science employ differing naming conventions, there may be confusion whenever considering common concepts, thus this thesis will endeavor to adhere to the naming conventions of graph theory more generally, and introduce the relevant network science terminology when referencing a concept unique to its domain.

degree of a node

$$k_i = \sum_{j=1}^N A_{i,j} \quad (2.2)$$

which in the undirected case represents the total number of connections, must be independently measured as either *in* (column sum) or *out* (row sum) degree.²

The powers of a non-normalized, unweighted adjacency matrix \mathbf{A} yield precisely the number of repeated possible walks from n_i to n_j . As a simple example of the above, we consider stepping from the second node of the pictured undirected, unweighted, non-looping simple graph. The adjacency matrix is constructed of the extant connections, i.e. the one-step paths available from node to node:



$$\mathbf{v}_2 \mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 3 & 1 & 1 \end{pmatrix} \quad \mathbf{A}^2 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$$

which yields the 2-step paths from n_2 to all other available nodes, as one may readily mentally verify through appreciation of the associated illustration above.³ Thus when considering the squared adjacency matrix, we have subsequently considered all transitions from every node $n_i \in N$ to every other node, and found their respective two step paths.

Weighted adjacency matrices \mathcal{A} may express non-binary relations between nodes, extending the definition of an edge from an ordered pair (i, j) to a triple $(i, j, w_{i,j}) \in \mathcal{E}$, where $w_{i,j} \in \mathbb{R}$ is an edge weight. Adding edge weight to directed networks allows for connections to be considered in the context of relative strength or frequency, which is essential when considering ecological, communication [36, 12, 18], transportation [32, 35], and social networks [15, 37, 23, 13], as many relations are not inherently reciprocal, as when a twitter user follows a celebrity, the celebrity is unlikely to follow the twitter user. Directed networks thus allow for expression of one-way channels, and when combined with edge weights, the relative strengths of all (including reciprocated) connections, as in [32] when the relative rates of passenger flux through airports is expressed by the network of airport connections by the total passengers on all flights between any two given airports. In this example, because the connections are given as sum total passengers over the course of a week's flights, the edge weight is affected by both the variations in quantity (number of passengers per flight) and frequency (number of flights between two airports over a week), which together constitute the conventional influences on edge weight.⁴ Whenever a network determines its edge weights in part by frequency, it implicitly assumes that its dynamics occur on timescales considerably greater than the frequencies reflected in its edge weights, as otherwise network adaptation would be based on a mercurial structure whose static representation at best represents an average over the time frame of the adaptation. [40]

Equipped with a minimal set of relations described by a network structure, one may apply a variety of different techniques to simulate network dynamics which parrot the empirical development a network

²This mathematical convention is computationally upheld through the python networkx library (specifically in numpy \mathbf{A} conversion) and a discussion of the effect of reversing these conventions and the resultant model directionality is considered in the Appendix, section 7.4.1

³There is no path stepping away from node 2 and ending on node 1, while stepping to nodes 3 or 4 both allow for either subsequently stepping to node 4 or 3, or else returning to node 2. Lastly, one may step to node 1 and return, thus making all possible paths from and then to node 2 the expected $(0 \ 3 \ 1 \ 1)$

⁴The fact that most networks map from interactions suggests that these interactions are averaged over some time interval which in general is considered to be small enough so as not to disturb the temporal analysis of network-wide behavior, however these assumptions have been challenged [39], and are briefly discussed in chapter 4

was derived from, or else designed to mimic. Some methods, such as preferential attachment as seen in the Barabási-Albert Model [36] or random rewiring as seen in the Watts-Strogatz Model [41], are simple enough to be used as minimal derivations for canonical models themselves, in this case scale-free and small-world networks, respectively. Other models of network adaptation, though generally far simpler than the empirical phenomena they describe, are nonetheless more complex, and may feature global properties (e.g. total population, edge weight) or metrics⁵ (e.g. average shortest path, routing efficiency⁶) in their attempts to describe the real-world behavior. (Consider how a twitter user's popularity may be implicitly affected by the total population and posting frequency of the entire twitter network) Grafting conventional physics onto that of a model has proven particularly fruitful, [42, 43, 44, 45, 46] due to the natural simplicity of many physical laws and their diversity of coupled behavior. As networks are often borne of a desire to understand the constituent nodes' interdependence, many models rely on diffusion⁷ as an analytic proxy for sequential communication, with e.g. news spreading through the network as nodes subsequently share or vote on stories. [18] The dynamics of diffusion may be described by random walkers, as per its conventional derivation through examining Brownian motion, and thus in the following we detail the analytic description of random walkers through a network for later use in an effective distance metric.

2.2 Random Walkers

In order to understand the notion of effective distance founded on random walks, we here develop a more formal description of the random walk. As a random walker's every step is taken without reference to their previous step, their walk is considered a *Markovian* process, evolving from state to state dependant only on their present position. An essential feature of random walks is the time intervals between their steps, and the probability distribution from which these time intervals are chosen. In the context of the Random Walker Effective Distance (RWED) however, we consider a constant time difference between all steps, which allows for more ready computation.⁸ Markov transition matrices may be constructed by normalizing the weighted adjacency matrix,

$$\mathbf{T}_{i,j} = \frac{\mathcal{A}_{i,j}}{\sum_{j=1}^N A_{i,j}} = \frac{\mathcal{A}_{i,j}}{k_i} \quad (2.3)$$

where the choice of normalizing by i or j determines if nodes possess normalized in or out degree. Building our notion of random walker processes from the base case of a simple, undirected graph, we first construct a transition matrix generally from the *weighted* adjacency matrix $\mathcal{A}_{i,j}$ as

$$\mathbf{T}_{i,j} = \frac{\mathcal{A}_{i,j}}{\sum_{j=1}^N \mathcal{A}_{i,j}} \quad (2.4)$$

as in equation 2.3, though with the caveat that for random walks, the row sums as defined in the denominator are normalized to reflect the transition probabilities directly;

$$\sum_{j=1}^N \mathcal{A}_{i,j} = 1 \quad \forall i \in \mathcal{N} \implies \mathbf{T}_{i,j} = \mathcal{A}_{i,j}$$

⁵A global network metric is one which relies on the entire network (e.g. network diameter), and subsequently any network dynamics contingent upon them implies every vertices' implicit awareness of all others

⁶For an introduction to global routing metrics, see section 4.3

⁷While some form of diffusion is most generally applicable directly, other more situation-specific effective distance metrics such as resistance distance [45] and random walker effective distance [32] will have the statistical mechanics of diffusion at their heart.

⁸Though the random walker effective distance makes gains in computational efficiency through making use of a constant time between (analytically) simulated steps, varying the probability distribution from which these waiting times are drawn, especially if this distribution should be coupled to nodes' edges, is a possibly fruitful avenue of study (as in [35]) not considered in this work.

Which makes the corresponding edge weights of an arbitrary node $i \in \mathcal{N}$ the probability that a random walker, finding themselves upon such a node, steps along a given edge, whose weight is thus given by the corresponding transition matrix/weighted adjacency matrix elements:

$$(i, j) \in \mathcal{E} \equiv \mathcal{A}_{i,j} = \mathbf{T}_{i,j} \leq 1$$

The introduction of normalized adjacency matrices allows for their interpretation as random walk Markov Chain transition matrices, where the i^{th} row $\mathbf{v}_i = \mathcal{A}_{i,:}$ may be considered a map from a vertex to the average of the neighboring nodes through the corresponding edges, $v_j \in \mathbf{v}_i$. If one begins on a given node $i \in \mathcal{N}$, and proceeds with equal probability along any of its edges, then for a given edge extending from node i to node j , the probability it would be selected is given by the j^{th} component of the vector obtained through multiplying \mathbf{v}_i by $\mathbf{T}_{i,j}$:

$$\tilde{\mathbf{v}}_k = \mathbf{v}_i \cdot \mathbf{T}_{i,j} \quad \equiv \quad \begin{pmatrix} & \tilde{\mathbf{v}}_k & \end{pmatrix} = \begin{pmatrix} & \mathbf{v}_i & \end{pmatrix} \begin{pmatrix} & \mathbf{T}_{i,j} & \end{pmatrix} \quad (2.5)$$

$\tilde{\mathbf{v}}_k$ is then the probability that at that latest step, a random walker on node n_i had transitioned to the n_j^{th} node, while $\mathbf{T}_{i,j}$ is thus the induced probability distribution of the node n_i , and by deriving $\mathbf{T}_{i,j}$ from a weighted adjacency matrix, one may ensure the network transitions consider the relative strength of network edges, when relevant empirical/a priori distributions are to be considered. [47] For an ergodic Markov process (i.e. a Markov chain with a finite state space) the *transition matrix* $\mathbf{M} \in \mathbb{R}^{n \times n}$ wherein $\mathbf{M}_{i,j}$ is the probability of transition between node n_i to node n_j may be given by the normalized weighted adjacency matrix, $\mathbf{T}_{i,j}$. Thus by multiplying $\mathbf{T}_{i,j}$ with itself one may find the probability of traveling from a given node n_i to its neighbors, and subsequently to the neighbors of its neighbors. Consider the case of transitioning from node $n_i \in \mathcal{N}$, where the initial step $\mathbf{v}_i \mathbf{T}$ yields the (average) position of the node n_i (i.e. given as a vector \mathbf{v}_i whose elements $\mathbf{v}_{i,j}$ represent the probability that node i steps to the j^{th} node.), and therefore subsequent steps, taking into account this initial transition, will be given as $\mathbf{v}_i (\mathbf{v}_i \mathbf{T})^m$, where m is the number of steps taken.

$$\text{Probability walker ends on node } n_j \text{ starting from node } n_i \text{ after } m \text{ steps} = \mathbf{v}_i (\mathbf{v}_i \mathbf{T})^m \quad (2.6)$$

In order to accommodate a history of the prospective random walker, we introduce time to the graph structure as in equation 2.8, and though this will assist in clarifying computation, the actual random walker processes will not be affected by the added structure, remaining a Markovian process.

$$\mathcal{G}_{t_i} = (\mathcal{E}, \mathcal{N}) \implies \mathcal{E} = \{(i, j, t = 0), (i, j, t = 1), \dots, (i, j, t = T)\} \forall i, j \in \mathcal{E} \quad (2.7)$$

As all timesteps are evenly spaced even for a potential ensemble of random walkers, timesteps reduce to step counts, and their interval is subject to interpretation according to the relevant comparisons to empirical phenomena. More generally, for dynamically adapting models which shift a network over time, it may be convenient to observe shifts via the change in adjacency matrix, which comprehensively describes the network. Thus, the dynamic network evolution may be seen as the set of networks at every sequential time step:

$$\mathcal{G} = \{\mathcal{G}_{t_0}, \mathcal{G}_{t_1}, \dots, \mathcal{G}_{t_{\text{final}}}\} \sim \mathcal{A} = \{\mathcal{A}(0), \mathcal{A}(1), \dots, \mathcal{A}(t_{\text{final}} \in \mathbb{N} = |\mathfrak{E}|)\} \quad (2.8)$$

When observing network dynamics, it is important to consider the relative rates of time sampling and network adaptation, [48, 35, 40] as without considering the rate of a network's structural evolution it is possible change is caused by forces acting on timescales smaller than those of the interval, and thus

leading to confounding, otherwise invisible shifts.⁹

For ready representation of edge values through time, we will suppress the node index in \mathbf{v}_i as the math is naturally not specific to any particular row, and subsequently any arbitrary row (or arbitrary node's outwardly directed edges) of sequential transition matrices $\mathfrak{T} = \{\mathbf{T}(0), \mathbf{T}(1), \dots, \mathbf{T}(t)\}$ after $t \in \mathbb{N}$ steps is given as $\mathbf{v}(t) \in \mathbb{R}^N$, where $v_i(t) \in \mathbf{v}(t)$ are the components, with edges directed from the i^{th} nodes. Making use of the transition matrix \mathbf{T} , we now work to define a random walk via $\mathbf{p}_i(t)$, the probability that a random walker (or ensemble thereof) rests on node i after $t \in \mathbb{N}$ steps. As the random walker must necessarily proceed somewhere after every step¹⁰ and end their walk on a node within the network regardless of the timestep t , probability conservation yields

$$\sum_{i=1}^N p_i(t) = 1 \quad \forall t \in \mathfrak{T} \quad (2.9)$$

where $\mathfrak{T} = \{t_0, t_1, \dots, t_{\text{final}}\}$ is the set of all time steps recorded in the model. To construct a dynamic equation for subsequent probabilities $\mathbf{p}(t+1)$ we use the same form from equation 2.5,

$$p_i(t+1) = \sum_{j=1}^N T_{i,j} p_j(t) \implies \mathbf{p}(t+1) = \mathbf{p}(t) \mathbf{T} \quad (2.10)$$

Now if we write out this form explicitly, starting from $t_0 = 0$, we find

$$\begin{aligned} \mathbf{p}(t_1) &= \mathbf{p}(t_0) \mathbf{T} \\ \mathbf{p}(t_2) &= \mathbf{p}(t_1) \mathbf{T} = (\mathbf{p}(t_0) \mathbf{T}) \mathbf{T} \\ \mathbf{p}(t_3) &= \mathbf{p}(t_2) \mathbf{T} = (\mathbf{p}(t_1) \mathbf{T}) \mathbf{T} = \mathbf{p}(t_0) \mathbf{T}^3 \\ &\vdots \\ \mathbf{p}(t) &= \mathbf{p}_{t-1} \mathbf{T} = \mathbf{p}(0) \mathbf{T}^t \end{aligned}$$

which hearkens to the earlier form seen in equation 2.5, where higher powers of the adjacency matrix \mathbf{A}^m returns the number of possible paths from one node to another, after m steps. As the transition matrix \mathbf{T} is row normalized and its values are less than or equal to one, $T_{i,j} \leq 1 \forall (i,j) \in \mathbf{T}$, powers of the transition matrix represents not the explicit number of different possible paths between nodes, but the probability that such paths are taken.

Knowing that the Markovian transition matrix $\mathbf{T}(t)$ describes the probability of arriving at a particular node after t steps, we may use this property to solve the eigenvalue equation and obtain the equilibrium distribution of \mathbf{T} :

$$\tilde{\mathbf{p}} = \lim_{t \rightarrow \infty} \mathbf{p}(t) = \bar{\mathbf{p}} \mathbf{T} \quad (2.11)$$

where $\tilde{\mathbf{p}}$ is the eigenvector of the eigenvalue $\lambda_0 = 1$, and $\mathbf{T}^{t_{\text{eq}}}$ is thus the probability distribution such that an ensemble of random walkers do not significantly alter their subsequent distribution among the nodes, $\tilde{\mathbf{p}}(t_m) = \tilde{\mathbf{p}}(t_{m+1}) \forall t_m > t_{\text{eq}} \in \mathfrak{T}$. Note that the tilde (as in $\tilde{\mathbf{p}}$) represents fulfillment of the equilibrium condition, i.e. $\tilde{\mathbf{p}} = \mathbf{p}(t > t_{\text{eq}})$. If we substitute the original set of N coupled equations for equation 2.10,

⁹As simulation of the Source to Structure model captures all shifts in network structure for every time step, this is only a necessary consideration in analysis and empirical comparison. Indeed, it is the purpose of the model for agents to be *reacting* to information being seeded, though care is taken in comparison to empirical data not to presume shifts in observables were due to structural shifts, and thusly beg the question. Furthermore, as network histories $\mathcal{A}(t)$ generated by the Source to Structure model have been shown to loose connections, but not regain them (As edges may be lost even in the case of uniform random edge initialization) and thus networks may be treated as static in analysis.

¹⁰Of course for graphs which allow for looping edges, it is possible for the walker at node i to remain in the same state/at the same edge with probability $\mathbf{T}_{i,i}$, though self-loops are not allowed in the Source to Structure model ($\mathbf{T}_{i,i} = 0 \forall i \in \mathcal{N}$)

we find that the row sums of \mathbf{T} are necessarily 1, as

$$\tilde{p}_i(t+1) = \sum_{j=1}^N T_{i,j} p_j(t) \implies \frac{\tilde{p}_i(t+1)}{\tilde{p}_i(t)} = \mathbf{1} = \sum_{j=1}^N T_{i,j}$$

where $\mathbf{1} \in \{1\}^N$ is here a vector of ones, and $\tilde{p}_i(t)$ is the equilibrium probability distribution of the i^{th} node at time t , though the time steps of $\tilde{\mathbf{p}}_i$ are only shown for comparison, as the value of $\tilde{\mathbf{p}}$ at a given timestep $t > t_{\text{eq}}$ are tautologically indistinguishable.

$$\tilde{p}_i = \frac{p_i(t_m)}{\sum_{j=1}^N \mathbf{T}_{i,j}}$$

As thermodynamics prescribes that a system in equilibrium is temporally reversible, we may consider the detailed balance of probability flow to and fro,

$$\tilde{p}_i(t) T_{i,j} = \tilde{p}_j(t) T_{j,i} \quad \forall t \in \mathbb{T} \quad (2.12)$$

which discussed in terms of directed networks and their corresponding adjacency matrix \mathbf{A} simply affirms that the sum total probability flow exiting nodes through their outwardly directed connections is equivalent to the overall inwardly directed flow. Seen in terms of row and column sums of \mathbf{A} this is especially clear,

$$\sum_i \sum_j \mathbf{A}_{i,j} = \sum_j \sum_i \mathbf{A}_{j,i}$$

and furthermore for row normalized transition matrices \mathbf{T} we know that the sum rows/columns will individually be one, and thus that the total sum will be equal to the number of nodes,

$$\sum_i \sum_j \mathbf{T} = \sum_j \sum_i \mathbf{T} = N$$

Making use of this normalization condition on \mathbf{T} , we can rewrite the detailed balance, without further assumptions, as

$$\begin{aligned} \tilde{p}_i(t) \sum_i T_{ij}(t) &= \tilde{p}_j(t) \sum_j T_{ji}(t) \quad \forall t \in \mathbb{T} \\ \tilde{p}_i &= \sum_j \tilde{p}_j T_{j,i} \\ &= \sum_j \tilde{p}_i T_{i,j} \\ \implies \tilde{\mathbf{p}} &= \tilde{\mathbf{p}} \mathbf{T} \end{aligned} \quad (2.13)$$

which shows that the solution to the detailed balance equation 2.12 yields the same probability distribution \mathbf{p} as the random walker after reaching a stationary state. After consideration, this is an expected result given that the random walker is a Markovian process whose stationary state is then both static in its distribution and whose behavior is inherently decoupled from earlier timesteps.

2.2.1 Continuous Time Random Walker

Converting to continuous random walks, we may predicate similar process with a *churn* or *flow rate* of γ , i.e. that the ensemble of random walkers transition to nodes every γ time units ($t_{n+1} - t_n = \gamma$), yielding

the probability for a node transition $n_i \rightarrow n_j$ as γT_{ij} . Let $p_i(t)$ assume its earlier role as the probability that the random walker is positioned on node i after t steps, which enables us to write the master equation:

$$\frac{dp_i(t)}{dt} = \gamma \sum_{j=1}^N [p_j(t)T_{j,i} - p_i(t)T_{i,j}]$$

Converting to matrix notation via the implicit imposition of δ_{ij} before the second term, $p_i(t)T_{i,j}$, this linear set of ordinary differential equations becomes

$$\frac{d\mathbf{p}(t)}{dt} = \gamma \mathbf{p}(t) [\mathbf{T} - \mathbb{1}] \quad (2.14)$$

Which is the form used in derivation of the analytic evaluation (discussed in the following section 2.3) of the random walker effective distance metric, [32] wherein the fundamental matrix [49] expressing the sum total expected visits (or ensemble probability) from the random walker after k steps as:

$$\text{Fundamental Matrix} = \sum_{t=1}^T \mathbf{T}^k = (\mathbb{1} - \mathbf{T})^{-1} \quad (2.15)$$

Equation 2.14 may be represented via the weighted degree, or diagonal matrix \mathbf{D}

$$\mathbf{D} = \begin{cases} \sum_j \mathcal{A}_{ij} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

and its Laplacian matrix \mathbf{L} ,

$$\mathbf{L}_{ij} = \begin{cases} \sum_j \mathcal{A}_{ij} & \text{if } i = j \\ -\mathcal{A}_{ij} & \text{otherwise} \end{cases} \quad (2.17)$$

as

$$\frac{d\mathbf{p}(t)}{dt} = -\gamma \mathbf{p}(t) \mathbf{D}^{-1} \mathbf{L} \quad (2.18)$$

Due to the normalization condition $\sum_{j=1}^N \mathcal{A}_{ij} = 1 = \sum_{i=1}^N p_i(t) = 1$, we may define the evolution with reference to the *random walk Laplacian* $\mathbf{L}' = \mathbf{D}^{-1} \mathbf{L}$, where we know that the diagonals of the Laplacian \mathbf{L} are 1. As the continuous random walker's choice of adjacent node and step timing are independent random processes, we deem the continuous random walker a *subordination process* [50], yielding the same dynamics in the discrete case¹¹. Specifically, if we consider the long-term equilibrium distribution, where $\frac{d\mathbf{p}(t)}{dt} = 0$, we find equation 2.14 becomes

$$0 = \tilde{\mathbf{p}}(\mathbf{T} - \mathbb{1}) \quad (2.19)$$

which is same as we know from the discrete process equilibrium equation (2.13).

2.3 Construction of Random Walker Effective Distance

In *Effective Distances for Epidemics Spreading on Complex Networks* [32], F. Iannelli et al.¹² advance the notion of effective distances formerly formulated in [37, 51] which exclusively relied on the *shortest path* between considered nodes¹³ to include *all paths* through an effective distance space defined fundamentally in [37] by

$$d_{nm} = (1 - \log P_{mn}) \geq 1 \quad (2.20)$$

¹¹The continuous random walker and the discrete random walker, ignoring the timestep γ , yield the same statistics.

¹²Especially Dr. Koher through his PhD thesis [35]

¹³Which in the empirical context of these original papers are airports and their respective passenger fluxes

where d_{nm} is the distance from node n to node m, and P_{nm} the normalized relative population fluxes. In the context of both original publications, a SIR model is considered as the basis of disease diffusion, and thus this is applied to three separate, dynamic, coupled populations¹⁴, however as [37] points out, and as in our less complicated case considering only one population, this is an unnecessary complication to the application of the effective distance metric. It should be noted that the log appears in order to reflect the necessarily additive nature of lengths and multiplicative nature of probabilities.

Here follows an abbreviated reconstruction of the random walker effective distance as in [32] and implemented in [52].

If we consider an ensemble of \mathfrak{N} random walkers, we may define the local transition rate \mathbf{Q} as the conditional probability that a randomly chosen node n_i transitions to another node n_j in the next time step,

$$\mathbb{P}\left(X_j^{(t+\delta t)}|X_i^{(t)}\right) \approx \mathbf{Q}_{ij}\delta t \quad i \neq j \quad (2.21)$$

distinct from the weighted, adjacency matrix \mathcal{A} in that the local transition rates \mathbf{Q} are normalized by the number of random walkers present on a given node (at a given timestep), $\mathfrak{N}_i(t)$.

$$\mathbf{Q}_{ij} = \frac{\mathcal{A}_{ij}}{\mathfrak{N}_i}$$

In this ensemble formulation, we have defined the adjacency matrix $\mathcal{A}(t)$ as the number of random walkers which have so transitioned $n_i \rightarrow n_j$ through timestep t . Both the local transition rates and the weighted adjacency matrix may be used to define the associated Markov transition matrix,

$$\mathbf{T}_{ij} = \frac{\mathcal{A}_{ij}}{\sum_j \mathcal{A}_{ij}} = \frac{\mathbf{Q}_{ij}}{\sum_j \mathbf{Q}_{ij}} \quad (2.22)$$

Though this explicit, discrete form of what has so far been conceived in the continuous limit of infinite random walkers, $\lim_{\mathfrak{N} \rightarrow \infty}$ may seem unwieldy in comparison, it highlights what would otherwise be invisible; that these terms; the local transition rates \mathbf{Q} , the weighted adjacency matrix \mathcal{A} , and Markov transition rates \mathbf{T} assume by default (and in application of these models to real data) different values. However, in the case of a continuous ensemble with normalized transition probabilities as implemented computationally in the Source to Structure model, we find that

$$\sum_j \mathcal{A}_{ij} = \sum_j \mathbf{Q}_{ij} = 1 \quad \Rightarrow \quad \mathbf{T} = \mathcal{A} = \mathbf{Q}$$

Initially considering only the shortest path effective distance as in [37] for conceptual clarity, the shortest path π_{ij} consists of n hops across edges $(k, l) \in \pi_{ij}$ where no node is visited more than once, an assumption which is eventually relaxed in the RWED. Our initial notion of an *effective distance* is then simply given as the sum of the inverted edges (as given by the weighted adjacency matrix) which constitute the shortest path from i to j .

$$\mathfrak{D}_{ij} = \min_{\pi_{ij}} \sum_{(k,l) \in \pi_{ij}} \frac{1}{\mathcal{A}_{kl}} \quad (2.23)$$

This inversion of the shortest path edge sum to form the effective distance is simply due to seeing the values of the weighted adjacency matrix as flux rates, which should naturally be *negatively* correlated with effective distance; the larger the flux (or in our interpretation, frequency/probability of passing through a given node) the *smaller* the effective distance. The shortest path effective distance measure implies a reliance on global information, as it is only via consideration of the entire graph that each agent may find

¹⁴The three distinct populations thus being susceptible (S), infected (I) and recovered (R)

their optimal, *shortest* path.¹⁵ The random walker effective distance does not assume global knowledge, and as it considers walks which potentially pass over the same nodes twice in their journey, they provide an upper limit to path length beyond even that of the multiple path effective distance [52, 32]. This discrepancy between the global and local routing efficiencies reflects notions of serial vs parallel processing methods both computationally and as a broader notion of how different processes are performed, and is discussed in greater detail in section 4.3.

Initializing our time-series with a proverbial walker initially at node i , assuming that the local transition rate is minimal $\mathbf{Q}_{i \rightarrow j} \ll 1$, (as is the case for $\lim_{\beta \rightarrow \infty}$) we may derive a Gumbel type probability density function to measure the extrema sampling, and in this case the first hitting time of a random walker transitioning from $n_i \rightarrow n_j$ is then given by its first moment,

$$\langle h_j \rangle_i = \frac{1}{\beta} \left(\ln \frac{\beta}{\mathbf{Q}_{ij}} - \gamma_e \right) \quad (2.24)$$

where $\gamma_e \approx 0.5772$ is the Euler-Mascheroni constant, and Q_{ij} is the local transition rate. This first moment is then re-arranged as

$$\begin{aligned} \beta \langle h_j \rangle_i &= \ln \frac{\beta}{\sum_j \mathbf{Q}_{ij}} - \gamma_e - \ln \frac{\mathbf{Q}_{ij}}{\sum_j \mathbf{Q}_{ij}} \\ &= \delta - \ln \mathbf{T}_{ij} \end{aligned} \quad (2.25)$$

Where $\delta = \ln \frac{\beta}{\sum_i \mathbf{Q}_{ij}} - \gamma_e$, and constant $\alpha = \sum_j \mathbf{Q}_{ij}$ are remnants of the SIR derivation, fixed at the value

$$\delta = \ln \left(\frac{\beta - \mu}{\alpha} \right) - \gamma_e \quad (2.26)$$

according to the transition rates $S \xrightarrow{\beta SI} I \xrightarrow{\mu I} R$ where β and μ are the infection and recovery rate respectively. However, in the context of the Source to Structure model, where the RWED is not bounded to keep the form of the SIR model¹⁶, δ is a parameter explored, and then fixed to allow for clarity in exploration of the remaining parameter space without any confounding effects.¹⁷ By minimizing this first hitting time $\langle h_l \rangle_k$ for each subsequent transition $(n_l \rightarrow n_k)$ in the path π_{ij} , we minimize the path, yielding the shortest path effective distance:

$$\mathfrak{D}_{ij}^{\text{SP}}(\delta) \equiv \min_{\pi_{ij}} \sum_{(k,j) \in \pi_{ij}} (\delta - \ln P_{kl}) \quad (2.27)$$

The shortest path algorithm is thus a special case of this more general effective distance measure, and the precise shortest path effective distance measure found in [37] as considered at the beginning of this derivation, may be found by setting $\delta = 1$ in the above equation 2.27. When considering the spread of information though a network other paths than the globally optimal, directed shortest paths must be taken into account to accommodate the potentially undirected (from the broadcaster's point of view) *diffusion* of information through a network. This kind of sequential spread through a network does not rely on the coordinated participation of those delivering the information, or in the context of much of this original research [32, 34, 37] viral pathogens, and thus the efficiency of a network in either directed

¹⁵The fastest time to compute the shortest path of a weighted directed graph so far discovered is in $\mathcal{O}(|E| + |V|\log|V|)$ time, via a modified version of Dijkstra's 1956 algorithm, where $|E|, |V|$ are the number of edges and vertices, respectively. [53]

¹⁶Indeed, it is precisely the natural behavior resulting from the network's dynamic adaptation which we intend to discover.

¹⁷A detailed exploration of graph-searches using various values of δ was performed, and results may be seen in the associated git repository. [54]

single-path routing, or diffusive, multiple simultaneous paths based routing suggests the degree of active agency or direction of the agents in the network, as discussed in section 4.3. There are multiple measures of this *multiple path* effective distance, including explicitly evaluating the efficiency of every possible path up to a given depth¹⁸ [35], though this rapidly grows computationally cumbersome, (combinatorically in the number of vertices, $\mathcal{O}(V!)$) as does a similarly explicit version as given in [31]. By expansion of the random walker effective distance to accommodate all possible paths however, Iannelli et al. [32] arrive at an analytic method which evaluates the long-time limit of an infinitely large, normalized ensemble random walker process through multiple paths.

Expanding equation 2.27 to allow for multiple paths from the same beginning and end nodes, yet retaining the notion that every path considered treads not upon itself, we arrive at paths so constructed of products of individual edge efficiencies, as in

$$\mathfrak{D}^{\pi_{ij}}(\delta) = \ln \left(\prod_{(k,l) \in \pi_{ij}} \frac{e^\delta}{\mathbf{T}_{kl}} \right) \quad (2.28)$$

This effective distance measure which accounts for the multiplicity of possible paths from source to target is able to combine the exponentiated distance between points as the simple sum of the exponentiated paths, as discovered in [34]:

$$e^{\mathfrak{D}_{ij}^{2T}} = e^{-\mathfrak{D}^{\pi_{ij}}} + e^{-\mathfrak{D}^{\pi'_{ij}}} \quad (2.29)$$

where π and π' are the two paths which are thus taken into consideration. Naturally, as this remains an exponentiated version of the effective distance, in order to reproduce the original effective distance metric, we take the natural logarithm of the result:

$$\mathfrak{D}_{ij}^{2T} = \ln \left(e^{-\mathfrak{D}^{\pi_{ij}}} + e^{-\mathfrak{D}^{\pi'_{ij}}} \right) \quad (2.30)$$

Extrapolating this form for accounting for multiple paths from two to arbitrarily many paths, one need only take the sum of the exponential, which, when combined with the need to take the log of the resultant exponentiated multiple path effective distance, yields the following form for the general Multiple Path Effective Distance (MPED):

$$\mathfrak{D}_{ij}^{MP}(\delta) = -\ln \left(\sum_{\pi_{ij}} e^{-n_{ij}} F_{ij}(\pi_{ij}) \right) \quad (2.31)$$

Where the probability associated with any given path π_{ij} is given by

$$F_{ij}(\pi_{ij}) = \prod_{(k,j) \in \pi_{ij}} \mathbf{T}_{kl} \quad (2.32)$$

where $|\pi_{ij}| = n_{ij}$ is thus the integer number of steps from node i to j , and summing over all path lengths yields:

$$\mathfrak{D}_{ij}^{MP}(\delta) = -\ln \left(\sum_{n=1}^{\max n_{ij}} e^{-n\delta} F_{ij}(n) \right) \quad (2.33)$$

Note that if only the dominant path $\tilde{\pi}_{ij}$ is chosen, we again arrive at the shortest path effective distance measure:

$$\mathfrak{D}_{ij}^{\tilde{M}P}(\delta) = \tilde{n}\delta - \ln F_{ij}(\tilde{n}) = \mathfrak{D}_{ij}^{SP}$$

¹⁸As is done in the *multiple path effective distance* algorithm as implemented Dr. Andreas Koher in his associated git repository, [52]

This describes the derivation of the *Multiple Path Effective Distance* (MPED), however as our model is based upon the *Random Walker Effective Distance* (RWED), we must make the jump to consider the Markov chain which describes paths which potentially intersect with themselves, retreading edges which are already part of the earlier path. This allows for the recursive definition of the path, as every path (or walker) no longer need possess a memory to eliminate potential (previously trodden) paths, yielding a RWED of \mathfrak{D}^{RW} :

$$\mathfrak{D}_{ij}^{\text{RW}}(\delta) = -\ln \left(\sum_{\{\Xi_{ij}\}} e^{-n_{ij} \delta} H_{ij}(\Xi_{ij}) \right) \quad (2.34)$$

where Ξ_{ij} are walks which unlike the paths π_{ij} considered in \mathfrak{D}^{SP} and \mathfrak{D}^{MP} allow for the walker to retrace his former paths, thus $H_{ij}(\Xi_{ij})$ is the probability associated with a walk that starts at i and arrives on j [32]. Grouping paths of equal length n , we find that the probability for a walk from i to j using n steps is given as the hitting time for a recursively defined Markov chain:

$$H_{ij}(n) = \sum_{|\Xi_{ij}|=n} H_{ij}(\Xi_{ij}) \xrightarrow{\text{recursively}} = \sum_{k \neq j} \mathbf{T}_{ik} H_{kj}(n) \quad (2.35)$$

If we consider the path length as a random variable with a probability distribution given by $H_{ij}(n)$, then its discrete Laplace transform $\tilde{H}_{ij}(c) = \sum_{n=0}^{\infty} H_{ij}(n) e^{cn}$ allows for the RWED to be defined purely in terms of a summation over path lengths n :

$$\mathfrak{D}_{ij}^{\text{RW}} = \ln \tilde{H}_{ij}(\delta) = \ln \sum_{n=0}^{\infty} H_{ij}(n) e^{-\delta n} \quad (2.36)$$

This form of the RWED, considered as an infinite sum over path lengths rather than over a combinatorially expanding number of paths hides the critical computational advantage: that this cumulant generating function over the hitting probability corresponds to the arrival time allows for the use of the analysis of Markov processes from random walk theory to determine a multiple *walk* effective distance, converging to $\mathfrak{D}_{ij}^{\text{RW}}$. Thus our goal, pulling from the derivation shown in Dr. Koher's thesis [35], is to adapt the present form of the RWED to be readily resolved through random walker-derived matrix manipulation.

We begin by considering the n^{th} power of the transition matrix \mathbf{T}_{ij} , which, as discussed in section 2.1, yields the probability that a random walker starting at node i arrives at node j after exactly n steps:

$$\mathbf{P}_{ij} = [\mathbf{T}^n]_{ij} \quad (2.37)$$

We may then combine this with our earlier derivation of the hitting time $H_{ij}(k)$ which as it predicts the first time a given node is visited, can abbreviate the exponentiation of the transition matrix by determining the probability that a node was visited in the first k steps, and then re-visits itself after $n - k$ steps:

$$\mathbf{P}_{ij}(n) = \sum_{k=0}^n H_{ij}(k) \mathbf{P}_{jj}(n-k) \quad (2.38)$$

Note that an essential feature of this definition is the idea that a random walker may, through the course of the n -step length walk, revisit the same node. By convention, we consider that for 0 length paths, $H_{ij}(0) = 0$, $P_{jj}(0) = 1$. Again making use of the discrete Laplace transform of the total probability over

all path lengths:

$$\begin{aligned}
 \sum_{n=0}^{\infty} \mathbf{P}_{ij}(n) e^{cn} &= \sum_{n=0}^{\infty} e^{cn} \sum_{k=1}^n H_{ij}(k) \mathbf{P}_{jj}(n-k) \\
 &= \sum_{k=0}^{\infty} H_{ij}(k) e^{ck} \sum_{k=n}^{\infty} \mathbf{P}_{jj}(n-k) e^{c(n-k)} \\
 &= \sum_{k=0}^{\infty} H_{ij}(k) e^{ck} \sum_{k'=0}^{\infty} \mathbf{P}_{jj}(k') e^{ck'}
 \end{aligned} \tag{2.39}$$

where we have swapped the order of summation and relabeled the index $(n - k) \rightarrow k'$, as convolution of probabilities turn to products in Laplace space. Collectively, these shifts allow us to write equation 2.38 as:

$$\left. \begin{aligned}
 \tilde{\mathbf{P}}_{ij}(c) &= \sum_{k=0}^{\infty} \mathbf{P}_{ij}(n) e^{ck} \\
 \tilde{H}_{ij}(c) &= \sum_{k=0}^{\infty} H_{ij}(n) e^{ck}
 \end{aligned} \right\} \quad \tilde{\mathbf{P}}_{ij}(c) = \tilde{H}_{ij}(c) \mathbf{P}_{jj}(c) \quad \Rightarrow \quad \tilde{H}_{ij}(c) = \frac{\tilde{\mathbf{P}}_{ij}(c)}{\tilde{\mathbf{P}}_{jj}(c)}$$

Having an explicit form for the Laplace transformed hitting time $\tilde{H}_{ij}(c = \delta)$ allows us to rewrite the RWED as defined in equation 2.36:

$$\mathfrak{D}_{ij}^{RW} = -\ln \tilde{H}_{ij}(\delta) = -\ln \frac{\tilde{\mathbf{P}}_{ij}(\delta)}{\tilde{\mathbf{P}}_{jj}(\delta)} \tag{2.40}$$

Using equation 2.37 we can rewrite the Laplace transformed sampling probability $\tilde{\mathbf{P}}(\delta)$ in terms of the Markov transition matrix \mathbf{T} :

$$\begin{aligned}
 \tilde{\mathbf{P}}(\delta) &= \sum_{n=0}^{\infty} \mathbf{T}^n e^{-\delta n} \\
 &= \sum_{n=0}^{\infty} [\mathbf{T} e^{-\delta}]^n
 \end{aligned} \tag{2.41}$$

where we deal only with values of $\delta > 0$, wherein the matrix $\mathbf{T} e^{-\delta}$ is sub-stochastic, and all row-sums are sub-zero,

$$\sum_j \mathbf{T}_{ij} e^{-\delta} \leq 1$$

as according to the conserved attention model row sums are normalized, $\sum_i \mathbf{T}_{ij} = 1 \forall i \in N$ and $e^{-\delta} < 1 \forall \delta > 0$. The dominant eigenvalue $\tilde{\lambda}$ of $\mathbf{T}_{ij} e^{-\delta}$ is less than 1, assuring that the limit of the geometric sum exists and is equal to the *fundamental matrix* [49]:

$$\tilde{\mathbf{P}}(\delta) = \mathbf{Z} = [\mathbb{1} - \mathbf{T} e^{-\delta}]^{-1} \tag{2.42}$$

which at last provides a method of computing the RWED via analytic matrix manipulation (in this case, the majority of the computation is spent in the inversion of a $\mathbb{R}^{N \times N}$ matrix, see section 7.3). In order to compute the complete RWED, we must then compute the logarithm of the quotient $\frac{\tilde{\mathbf{P}}_{ij}(\delta)}{\tilde{\mathbf{P}}_{jj}(\delta)}$, where the numerator is exactly the fundamental matrix, and the denominator may be given by the inversion of the diagonals of the fundamental matrix:

$$\tilde{\mathbf{P}}_{jj}(\delta) = \text{Diag}(\mathbf{Z})^{-1}$$

where the Diag simply takes the diagonal elements of a given matrix, which makes its inversion particularly easy (simply directly inverting all individual remaining diagonal elements) which yields the final form used to compute the RWED in the Source to Structure model as:

$$\mathfrak{D}^{RW} = -\ln \left[\mathbf{Z} \cdot \text{Diag}(\mathbf{Z}^{-1}) \right] \tag{2.43}$$

This final form may be computed in computational complexity of $\mathcal{O}(N^{2.4})$, which though slightly exceeding that of the shortest path effective distance which requires merely single iteration of Dijkstra's algorithm in $\mathcal{O}(N^2 \log N)$ time [55, 53], the RWED compensates with consideration of all possible paths from target to source.

Chapter 3

Model Description

Here follows a description our model for a local network adapataion process inspired by recent epidemiological [32, 37, 12] and sociophysics research [13, 19, 15, 23] and closely resembling in intent and form the models used in [14]¹, the python implementation of which may be found at the associated github repository [54].

3.1 General Model Form

Most generally, the Meme Migration Model consists of seeding information onto a network and subsequently letting agents (nodes) adjust their connections to their neighbors depending on the neighbor's proximity to the source. More specifically, a simple², directed³, weighted⁴ network is constructed and information is seeded onto the network, choosing a node or nodes to be the *source*, whereafter diffusion is pseudo-simulated through an effective distance metric via edges extending from the source.⁵ All agents then reinforce their incoming edges proportional to the information these edges received from their source node in this latest step, as a fraction of the total information the agent received from all its edges. This ensures that all agents/nodes reinforce their locally optimal connections while also maintaining the critically conserved quantity of attention (incoming nodes), as all incoming edges from each node are normalized after every diffusion iteration. The constant re-normalization of every agent's total incoming edge values ensures that edges which perform worse than average are penalized, allowing for some connections to fade (be normalized) virtually out of existence. Though by default we may consider all nodes' outgoing edges normalized to the same value, by allowing for a distribution of how the nodes are normalized, one may allow for varying levels of *information projection capacity*, and furthermore via modification of all parameters/binary choices we discover a variety of different behaviors which will be discussed at length in section 5.

Without further considerations, this most basic formulation of network diffusion allows for considerable choice; in initial network structure, in diffusion and seeding mechanisms, in diffusion and adaptation directionality in the rates of adaptation and information dispersal, and it is the latter we will consider in hopes that via variations in the rates and mechanism of diffusion we may discover network adaptation which leads to different network structures. The only defining feature of this most basic formulation is that for every run, information diffuses from a minimum of one node in the network, and the network

¹Despite the similarities in model form, the model was developed independent from Almaatouq et al.

²A simple network is one without vertices connecting an edge to itself (i.e. loops)

³Wherein every connection contained information about directionality; e.g. node A to node B (but not B to A)

⁴Weights imply that the connections between vertices (edges) are non-binary

⁵Note that we consider the diffusion of information as a most general, yet immediately relevant case in the context of rising social media usage, however in its most abstract form this diffusion process may apply to anything from the spread of disease [32, 33] to the adoption of a behavior or opinion through socialization [56, 15].

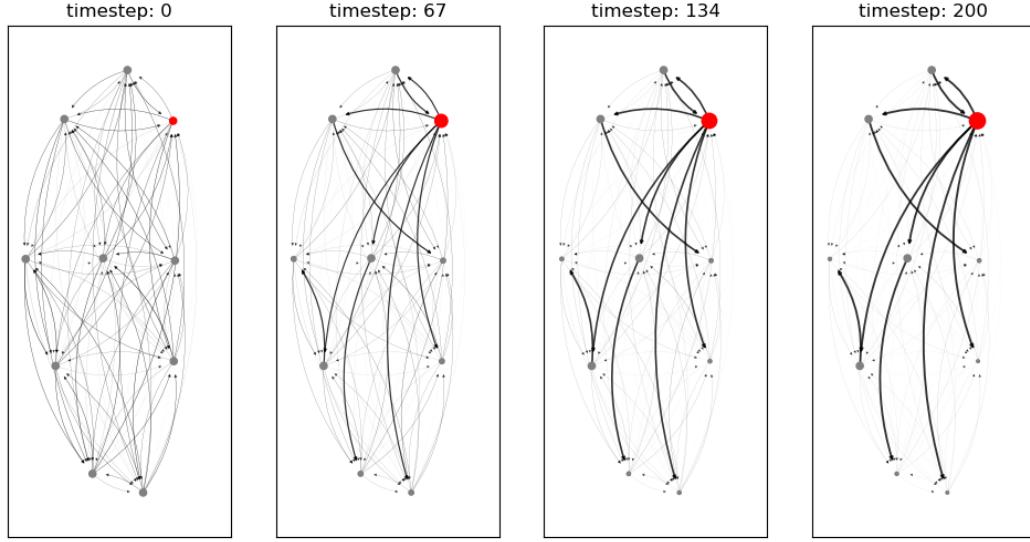


Figure 3.1: Evolution of a simple, directed network’s edge weights due to diffusion. The red node labels the source, which is set constant in this test case to demonstrate the network’s adaptation in the base case of constant seeding. This network picture results from use of the multi path effective distance algorithm, with a *take the best* edge reweighting algorithm.

adjusts its edges in response. The immediately relevant variables are: choice of how (and thus where, how much per round) to seed information into the network, the diffusion process (how readily can secondary, tertiary paths, e.g. friends of friends, deliver information⁶), and naturally the size of the network.

3.2 Network Initialization

There are various methods of initializing the requisite simple, non-looping, connected, directed, weighted network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, which in practice consists of constructing the corresponding adjacency matrix, \mathbf{A} . One may simply use known empirical networks or scaled versions thereof, however for purposes of general exploration using the fewest assumptions of network structure possible, the network was initialized into several well known edge structures from network science, each naturally accommodating arbitrary network size: uniformly random, scale-free with a set degree exponent γ_{sf} , or sparsely with a set in/out node degree, optionally equalizing initial edge values.

Uniform random edge initialization predictably assigns edges values from a uniform random probability distribution, before normalizing node out degree⁷. Scale free edge initialization assigns every node at least a predefined minimum number of distinct edges k_{\min} , and otherwise the number of connections per node is drawn from the power law distribution $p_k \approx k^{-\gamma_{sf}}$, with \mathbf{A} afterwards row normalized. Sparse edge initialization simply ensures every node possesses the same number of randomly chosen edges with equal weight, and subsequently normalizes outgoing edges.

For both sparse and scale-free edge initialization, care is taken to ensure that the network remains connected, so that there are no nodes without another node connected to them, or mathematically speaking, $\nexists i \in |\mathcal{N}| s.t. \sum_i \mathbf{A}_{i;j} = 0$, i.e. all columns have non-zero sums. Furthermore, all network initializations are

⁶Generally, these multi-step paths will hereafter be considered as *higher order paths*

⁷Node out degree is conventionally given as the rows of the adjacency matrix, thus normalization is performed as $\mathbf{A} \rightarrow \frac{\mathbf{A}}{\sum_j \mathbf{A}_{i;j}}$

optionally undirected, whereby reciprocal connections are averaged, i.e. $e_{ij}, e_{ji} \rightarrow \frac{e_{ij} + e_{ji}}{2} \forall e_{ij} \in \mathcal{E}$, and such a scheme of 'undirectification' via continuous averaging after edge adaptation is available through the entire run time.⁸

For all base tests, the network was initialized via a uniform random all to all distribution, which allowed for exploration of all base cases while providing a sufficiently stochastic substrate for sub-strategies determined by initial differences to show their effects.

3.3 Model Synopsis

Before giving a more detailed description of the Source to Structure model, a simplified summary is here detailed for ready reference. Following the initialization of a simple, directed, connected, weighted graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ with either a uniformly random, sparse, or scale-free structure (which consequently determines the form of the adjacency matrix \mathbf{A}), the model adapts through the following process:

1. A source node $n_s \in \mathcal{N}$ is selected via a map $S_{\text{Seeder}}(\mathbb{R}) : \mathbb{N} \rightarrow \mathcal{N} \in \{S_{\text{pwr law}}, S_{\text{shifting}}, S_{\text{constant}}, S_{\text{normal}}\}$ as discussed in section 3.4.
2. The Random Walker Effective Distance (RWED) from the source node s to all other nodes $n_j \in \mathcal{N}$ is evaluated (section 2.3), producing $\mathfrak{D}_{s;j}$:

$$\mathfrak{D}^{\text{RW}} = -\ln \left[\mathbf{Z} \otimes \text{Diag}(\mathbf{Z})^{-1} \right] \quad \text{where} \quad \mathbf{Z} = \left[\mathbf{1} - \mathbf{A} e^{-\delta} \right]^{-1}$$

3. The RWED from source to itself (naturally always 0) is replaced with a fraction of the lowest ED:

$$\mathfrak{D}_{s;s} = \frac{\min(\mathfrak{D}_{s;j \neq s})}{2.6}$$

4. From \mathfrak{D} and \mathbf{A} every edge's info-scores (section 3.6) are evaluated according to $\mathcal{S}(\mathbb{R}) : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{N \times N}$:

$$\mathcal{S}_{i,k} = \left(\frac{\mathfrak{D}_{ik}}{\sum_j \mathfrak{D}_{i;j}} \right)^{\kappa-1} \cdot \mathbf{A}_{ik}^\kappa$$

5. For every node, reward the top ϕ percentile of their edges, as according to their info scores (section 3.6.2):

$$\frac{d}{dt} \mathcal{E}_{i,j} = \begin{cases} > 0, & \text{if } \mathcal{S}_{\text{final}}(i, j) > 0 \\ \leq 0, & \text{otherwise} \end{cases} \begin{cases} \mathcal{S}_{\text{final}}(i, j) & \text{if coupled } \mathcal{S} \\ 1 & \text{else} \end{cases}$$

and subsequently normalize all the node's lately adjusted edges, reducing the value of unrewarded edges.

6. Perform steps 1 through 5 at least once per run, and observe the evolution of network edges $\mathbf{E}(t)$ through time.

⁸Though naturally observing the behavior of the model wherein all initial connections are reciprocated and whereafter they are forced to be produce differing results, discussed in the appendix 7.4.1

3.4 Information Seeding

As a null case without any coupling between the structure of the network and its information dispersal, we consider *constant* and *random seeding*, wherein for every diffusion step⁹ information is seeded onto either the same or a new, randomly selected node, respectively. These methods provide for knowledge of the model behavior in base cases, wherein we wish to consider and understand the complexity of the diffusion process with minimal assumptions and corresponding complications. In order to compare to real systems, we may also consider seeding based on an a priori distribution, or even based upon the dynamic network structure itself, e.g. based on every node's diversity of connections¹⁰ Furthermore, as an intermediate between constant and random seeding, one may consider running the network with a set number of switches of the constant source, such that during the course of a model's run-time the source changes randomly to nodes which have not yet become the source¹¹.

More formally, random seeding simply selects a new node at random for every diffusion step, or after a set number of diffusion steps, while constant source seeding naturally never shifts the source node through the course of the simulation. Slightly less trivial than the former seeding scenarios, power law and normally distributed seeding both choose a new seed for every step from a power law $P(x; a) = ax^{a-1}, 0 \leq x \leq 1, a > 0$, or normal $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ distribution respectively. In both power law and normally distributed seeding, the network has more opportunity to adapt to optimize for the relatively consistent source selection (in comparison to those methods which select new sources at random), behavior seen in empirical systems, such as in social media networks [39, 57]. Seeding functions thus take the form $S_{\text{Seeder}}(\mathbb{R}) : \mathbb{N} \rightarrow \mathcal{N}$ where the optional parameter \mathbb{R} may specify the standard deviation σ for normally distributed seeding, $S_{\text{normal}}(\sigma, \mu \in \mathbb{R})$, the degree exponent for $S_{\text{pwr law}}(a \in \mathbb{R})$, or source for S_{constant} . Regardless, every seeding function prescribes to every time step, denoted $t \in \mathbb{N}$, a corresponding source $s \in \mathcal{N}$.

Every case of seeding choice represents a different source of information into the network, which is understood to be a function of the greater system into which our representative model is itself embedded. For instance, the notion of constant seeding is consistent with a single news source [58], while that of normal or power-law distributed seeding potentially portrays the effect of overlapping communities, wherein the most centrally connected in another community therefore acts as a common source of information in the modeled network. These a priori seeding distributions therefore require another assumption or comparison to known empirical information source distributions¹² such as how social media is known to produce power-law activity frequencies [39, 57]. The requirement may be slightly alleviated via a seeding mechanism which is itself dependant upon the network, and simply observing the resultant behavior, as this implies a significance to the coupled seeding mechanism and the network adaptation. Unfortunately, this added complexity (often including the addition of another variable) may be more obfuscating than elucidating, as regardless of the efficacy of the match to known results from an entirely autonomous system, the coupled response to single parameter shifts conceals the relevant connections. This stated, should all other parameters be fully tested against uncoupled source seeding, it is possible to consider coupled seeding mechanisms with some hope of recovering the effects of individual parameters. We endeavor to perform such analysis controlling for the effects of all parameters individually, as discussed in section 5.2 and catalogued more comprehensively in the appendix.

⁹ As distinguished from a run in that every step necessarily consists of only one instance of source selection, i.e. information being seeded onto, and diffused through, the network, whereas a run may contain a set number, potentially > 1 of these diffusion steps

¹⁰ Which would naturally incentivize all nodes to fully weight their incoming edge with the same node, who would then (having the greatest diversity of connections) receive the most information directly. Implementation of this seeding method is discussed in Appendix: Information Seeding.

¹¹ Naturally, if the number of source shifts exceeds the number of nodes, there will necessarily be some repeat source nodes

¹² Known source distributions still require the assumption that the model's depicted network overlaps with that of the known distribution

3.5 Information Dispersal

As edges are the only feature of the network which persist, sustaining only partial modification after every run, information diffusion is evaluated with reference to their values, and they are then adjusted according to this latest evaluation. In order to avoid the difficulties associated with a naïve notion of direct diffusion, as described in Appendix: 7.4.2, we make use of effective distances, a generalized notion of distance initially developed through the mid twentieth century [38] mapping economic, technical or other constraints to better acknowledge the practical distances in their relevant fields. More recently, effective distance measures have gained favor among epidemiologists and network scientists in attempts to better depict the natural diffusion of disease and ideas in the context of airports and instant messaging [34, 32, 59]. Fundamentally, the hope is to find a consistent transformation into effective distance space whereafter the relevant behavior may be described via known physics, as performed for virus outbreaks in [37], which found viral infections to *diffuse* through their effective distance space. An effective distance metric also allows for consistent evaluation of information diffusion, or an unchanging notion of *distance from source*, compatible with global parameter shifts and higher order paths, unlike simple diffusion processes which may rely on artificial constraints, e.g. initial choice of path, whether or not information is allowed to revisit nodes, &c. Furthermore, the random walker effective distance (RWED) metric, as formulated in [32], originally derived in [35] and explained in section 2.3 allows for analytic evaluation, greatly speeding computation and allowing much larger networks.

Derived in section 2.3, the RWED is a map $\text{RWED} : \mathcal{E} \rightarrow \mathbb{R}^{N \times N}$ from the set of edges $\mathcal{E} \subset \mathbb{R}^{N \times N}$ of \mathcal{G} to their effective distances from each other, given as $\mathfrak{D}_{ij} \subset \mathbb{R}^{N \times N}$ such that the distance from source $i \in \mathbb{R}^N$ to target j is the i, j^{th} element of \mathfrak{D} . As only a single node is considered the source for a given time step, only one column containing the effective distances from chosen source to all other nodes is considered per seeding step. The complete process of evaluating effective distance for a given source node is as follows:

1. Row normalize the adjacency matrix: $\mathbf{A} \rightarrow \bar{\mathbf{A}} = \frac{\mathbf{A}}{\sum_j \mathbf{A}_{i,j}}$
 2. Perform RWED calculation: $\text{RWED}(\bar{\mathbf{A}}) \rightarrow \mathfrak{D}_{ij}$
- * Substitute the effective distance of the source itself with a value smaller¹³ than the minimum remaining set.

$$\mathfrak{D}_{i;j=i} = \frac{\min(\mathfrak{D}_{i;j \neq i})}{2.6}$$

Thus in this formalism, information is less *dispersed* through the network as its access is simply evaluated on a node-by-node level, a function of each node's direct and indirect (that is, through higher order multistep paths) connection to the source node.¹⁴ This evaluation relies upon the RWED and thus the result is the probability that a random walker starting on the source would step directly onto a given node, as given by the outwardly directed edge values of the source node.¹⁵ The literal interpretation of the RWED is however less enlightening than considering it as the proxy for spatial or temporal distance it is intended to represent; more intuitively, it may be considered as an arrival time for information/disease. Indeed, this is precisely how it is both derived (see section 2.3) and what it represents explicitly in [32, 35]. The indirect evaluation of effective distance to the source remains relevant, as the random walker

¹³By default, we divide the min of the remaining effective distance values by 2.6

¹⁴For ease of reference, when describing the model we will refer to the source node in the singular, though there may theoretically be many source nodes per run

¹⁵The default case choosing to consider effective distance *from* the source node (i.e. along *outwardly* directed edge from the source) is a convention chosen to simplify discussion; a discussion and examination wherein these default directionality choices are reversed is covered in section 3.7

effective distance does not simply consider the outward connections of the source, but (analytically) simulates an exhaustive series of random walks in which the higher order connections (i.e. secondary, tertiary connections to source wherein the walker must make several steps from the source to reach its target), if strong, are likewise likely to be visited, and thus may attain lower effective distance than nodes whose direct connection to the source may be otherwise stronger, but lacks in secondary connections.

Random walk distance broadly considers the number of steps it should take a random walker, proceeding with each step from node to node across their (normalized, incoming) edges, with the probability of taking the next step along a given path equal to its weight, to reach a given node when starting at the seeded node. This method has the advantage of an analytic solution (equivalent to exhaustively running random walks from a given seed position, as explained in section 2.3) which greatly reduces computation, as well as providing an intuitive solution. The multiple path effective distance algorithm yields an effective distance based on every possible path information may take from the source to the target node, (where every node is considered as a target node independently each step, yielding an effective distance for every node) and the value of higher order paths (those that require increasingly more than one hop to connect the nodes) may be adjusted by changing the longest allowed path length.¹⁶ This final evaluation of effective distance from source to all nodes is then used in edge evaluation, which however requires a further calculation of *info scores*, a kind of meta-effective distance metric which allows for intermediary edge evaluation for nodes whose dependence on global parameters provides insight into the effect of different presumed psychological heuristics, set on continuous scales. Details thereupon are provided in section 3.6

The notion of relative distance in terms of information diffusion may be more readily formulated in terms of the frequency contact (which we consider as weights of the edges). For instance, if a cat meets a dog every week, yet they have a mutual badger friend whom they meet independently every day, then it should be far easier for the cat to pass information to the dog through the badger and vis-versa than directly through their weekly encounter¹⁷. Furthermore, the exponential decline of information relevance cannot be readily shifted to reflect the ease or difficulty of indirect communication/diffusion, as should necessarily be a test parameter. We therefore turn to more recently tested methods of diffusion as considered in [37, 32, 34] and published with a public python implementation by Dr. Koher.

3.6 Network Adaptation

After information has diffused through the network, the network adapts by adjusting its edges' weights while always conserving each node's total incoming edges' weights. This step takes place every run, after information has been dispersed through the network, and is then adapted according to every connection's *info score*, an intermediary score designed to reflect the efficacy of a connection in the context of psychological heuristics. The info score function $f : \mathfrak{D} \subset \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^N$ is a function of both an edge's standing value and its effective distance (as a fraction of the target node's total edge effective distance values) from the source. Though the info score function may take several forms depending on the heuristic presumed to be at play, the following allows for an interpretation of a system parameter κ which tunes between pure dependence on the effective distance to the source node at $\kappa = 0$, and pure dependence on the standing edge value at $\kappa = 1$, with a smooth transition between:

$$\text{Info Score}_{ik}(\kappa) = \left(\frac{\mathfrak{D}_{ik}}{\sum_j \mathfrak{D}_{ij}} \right)^{\kappa-1} \cdot \mathbf{A}_{ik}^\kappa \equiv \mathcal{S}_{ik} \quad (3.1)$$

¹⁶Higher order paths are also generally cutoff at 2, to save on computation time, though this makes a surprisingly little difference; longer paths are rarely as influential due to the natural exponential suppression of their information's value as it percolates through the network

¹⁷This analogy assumes all meetings are bidirectional, i.e. undirected connections, something patently false for many networks. (Consider a twitter network, where follower/followed counts follow an exponential relation [60])

where κ is therefore the *edge conservation* parameter, $\mathfrak{D}_{i;j}$ the i^{th} node's effective distance (from the source node n_j for the directed connection), and $\mathbf{A}_{i;k}$ are the edge values connecting to the i^{th} node from every one of its connections, which are evaluated separately.

The first component, $\left(\frac{\mathfrak{D}_{ik}}{\sum_j \mathfrak{D}_{ij}}\right)$ yields the probability that a node receives information from a given node n_i *first*, while the second component \mathbf{A}_{ik} loosely describes the agent's willingness to believe the information, and κ modifies the emphasis on either trait. A high κ will thus lead to a reluctance to shift one's connections in light of new information, while a low κ leaves the agent immediately adapting to whatever the new source is to the best of its ability given its extant connections. As shown in chapter 5, in the present implementation higher values of κ have little benefit for local optimization, however account for the limited speed of adaptation seen in many real world systems, and is thus included as a potential reference.

The information received from a given edge is then given as the info score of the edge value and the node from which it is directed. Every node only *positively* adjusts the connections through which they have received information, i.e. have a nonzero info score, e.g. for node $i \in \mathcal{N}$, only edges $k \in \mathbb{A}_i : \mathcal{S}_{i,k} > 0$ are rewarded, generally proportional to the information they have received through each connection. Note that renormalization of edge values will necessarily reduce those edges which were not rewarded in a given run, as after normalization they are reduced proportional to the gain of other edges the node possesses. Furthermore, normalization is necessarily performed after all edges have been updated such that normalization does not interfere with info score evaluation.

3.6.1 Reinforcement Distribution

Not all edges are however reweighed after every run; the system's *selectivity* ϕ is used to eliminate the lower ϕ fraction of edges from the reinforcement pool, with their relevant relative rankings determined by their info scores. The remaining $1 - \phi$ fraction of edges¹⁸ redistribute their weight either proportionally to that node's remaining edge's info scores, or else simply reward all remaining edges with the same constant value, as depicted in figure 3.2. If we define the sorted set of info scores for a given node i as $[\mathcal{S}(i, l)] = \{l \in \mathcal{E}_i | \mathcal{S}(i, l - 1) < \mathcal{S}(i, l) \forall l \neq 0\}$, then we may consider the final set of info scores $\mathcal{S}_{\text{final}}(i, k)$ used to reinforce a node's edges as,

$$\mathcal{S}_{\text{final}}(i, k) = \begin{cases} \mathcal{S}(i, k), & \text{if } k \in \left\{ k \in [\phi \cdot [\mathcal{S}(i, k)]] \right\} \\ 0, & \text{otherwise} \end{cases} \quad (3.2)$$

The final set of final info scores $\mathcal{S}_{\text{final}}$ used to determine absolute or relative rewards¹⁹ thus determines the *selectivity* of the nodes in their choice of which connections to reinforce, with high ϕ translating into nodes only reinforcing their best performing connections. This parameter ϕ thus serves as another system parameter which is tested via grid-search, and indicates the threshold²⁰ which edges must pass before the node in question will reinforce them. The node will not necessarily reinforce the edges with the lowest effective distances (to the source node), but those with the highest info scores, as determined by equation 3.1.

Info scores represent a range of psychological heuristics in their dynamic definition via κ ; a high κ ($\kappa \approx 1$) represents a high dependence on preexisting edge values (thus called edge conservation), while a small κ value represents a high dependence on effective distance. As the effective distance represents the relevant optimization parameter throughout the simulation, it would appear that the best strategy would always emerge under info scores with low κ values, however the notion that agents should be able to both

¹⁸whose info scores are thus in the $1 - \phi$ percentile of all the node's edge's info scores

¹⁹The distinction between rewarding an edge simply if retains a positive non-zero value in the final info score ($k \in \mathcal{S}_{i,k} \subset \mathbb{R}_{>0}$), and rewarding the edge proportional to its final info-score is a system level choice discussed in appendix, 7.4.1

²⁰Relative to the node's other incoming edges

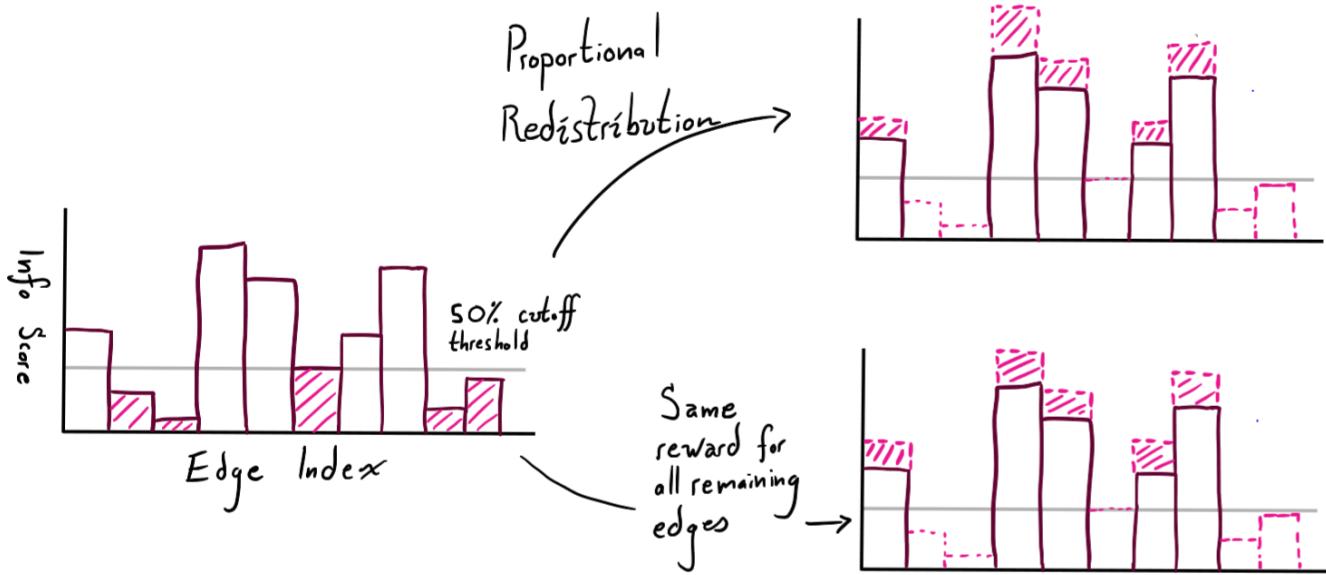


Figure 3.2: Reinforcement redistribution based on cut-off threshold ϕ , as a fraction of total edges, not max info score values. Note that the reinforcement is added to the edges values, *distinct from their info score* which however has a positive correlation thereto (lest for $\kappa > 0$)

immediately identify and optimize for this result seems forced at best; thus as κ increases beyond 0 it comes to represent the natural uncertainty of the agents in their adaptation, as well as a simple reluctance or transaction cost of immediately purging their existing contacts in favor of what they know to be optimal.

3.6.2 Edge Reweighting

The final form of edge reweighing used in the Source to Structure model is as above, *via the info score*, though not necessarily proportional to it. Through the use of info score to select the edges which are to be reinforced, the sensitivity to effective distance or extant edge value is controlled via the corresponding coupling parameter κ and thus needs not bleed into edge reweighing method as it otherwise might, which is in turn determined by the parameter ϕ . These methods of edge reweighing, and thus network adaptation,

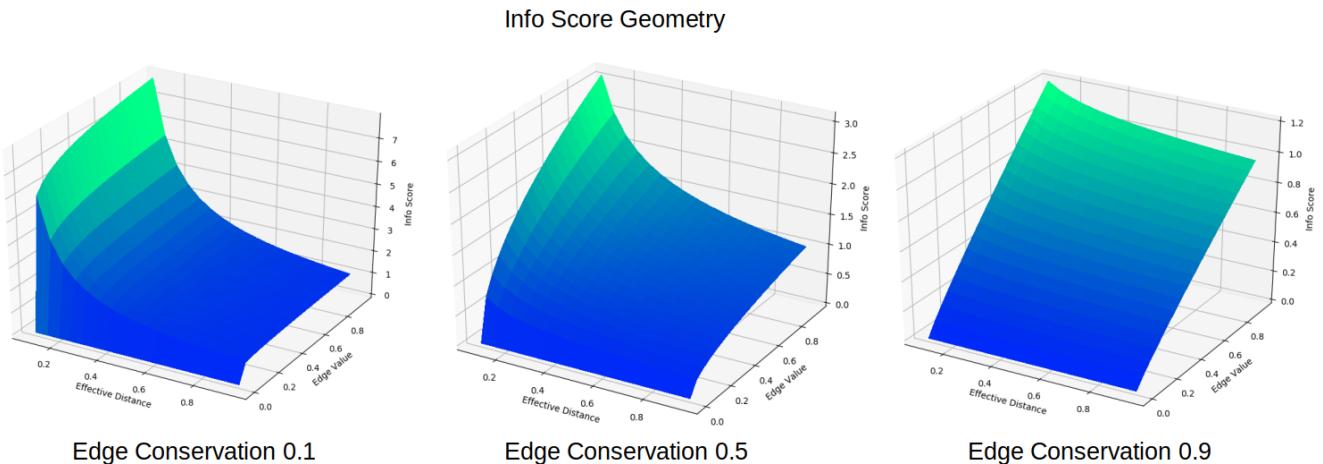


Figure 3.3: The geometry of info score space, or values of κ from left to right of $\kappa = 0.1, 0.5, 0.9$

in our model are chosen to test a wide variety of psychological heuristics which are thus theoretically

independent of the network dynamics. In this interpretation, the choice of ϕ is a proxy for agents' exclusivity in their adaptation; how many of their connections do they allow to influence their choice? If ϕ is low, representing a low fraction of connections which are dismissed, then the agent will only adapt as their network does; though they understand that some of their connections are considerably more effective at receiving and distributing information, they remain open to many of their other contacts, presumably in the hopes that these lately irrelevant contacts will soon become far more relevant (under change of source). A high ϕ leads to a more concentrated set of contacts which are considered, and this may lead to inefficiency under change of information source, as the relevancy of their former contacts is erased, and they must slowly adjust to the new information paradigm. However, while information continues to flow from the same source or cohort of size $N \cdot (1 - \phi)$, this method is far more efficient, as it requires the maintenance of far fewer connections²¹. It should also be noted that so long as the source does not change greatly (remains within the same highly connected cohort) the agent with a high ϕ value should not greatly suffer in efficiency.

3.7 Model Directionality

As the model operates upon a directed network, the choice of directionality is significant; in effective distance calculation (*to* or *from* the source), in edge adaptation (nodes adjusting their *outwardly* or *inwardly* directed edges), and in edge conservation (whether a node's *outwardly* or *inwardly* directed edges are conserved, and thus renormalized, after every run). For every one of these binary choices, there are coupled interpretations, as when all are either in the *from* or *to* configuration, that is, the network is operating on either following information *from* the source and adjusting (conserved) incoming edges, or else following all nodes' paths *to* the source, there is no difference as all their coupled effects counteract each other, and the resulting behavior is the same.²² A full discussion of implemented directionality shifts is considered in section 7.4.1.

²¹Here the maintenance of connections is understood to *cost* agents because all agent's connections are normalized, which means that anything spent in the preservation of one connection is at the cost of another.

²²This belies the arbitrary choice in computational interpretation in the reflected directionality of row or columns as edges pointing outwards or inwards.

Chapter 4

Network Observables

4.1 Observables

Though classification of network efficiency and hierarchy via the morphospace defined in section 4.2 provides important context for both the expected performance and evolution of the network structures produced through the Source to Structure model, other, more conventional observables from network science are necessary to better quantify their performance more broadly.¹ While there are a large number of metrics developed over the last fifty, and especially twenty, years for network analysis which may now be found in any current introductory network science text such as [25, 36], only a small set of computationally feasible and individually insightful observables were used to understand the Source to Structure model's network evolution. The set of relevant observables are further reduced by the Source to Structure model's conservation of (either in or out) degree, ensuring one of the network's directed degree distributions, and related metrics, remain constant though the course of the network's evolution.

4.1.1 Effective Distance Metrics

As the actual random walker effective distance was introduced in section 2.3, here follows a description of the relevant meta-metrics derived from the random walker effective distance. Effective distance from the source to all nodes, or else from all nodes to every other node (deemed all to all or alternately global, effective distance) is used as the primary metric representing gains in communication efficiency between the source and all other nodes, or between any given pair of nodes, respectively. Naturally the network's ability to adapt to a consistent, or even perfectly constant source would be considerably improved over the network's effective distance from source performance when the source is inconsistent and the network is not given time to adapt, though this naturally depends on the rate of network adaptation, which is investigated independently in section 5.1.1.

When considering grid-search results, it is necessary to have a single number for every parameter combination, and to that end we make use of *averaged* effective distance *differences*, wherein the initial effective distance (either from source or all-to-all) is subtracted from its respective final effective distance value. This difference, though especially variable in the context of effective distance to source and varying source², provides a picture of the performance of the system's evolution; if positive, then this reflects a larger effective distance than initially, and thus a poorer performance.

Note that both all to all and source to nodes effective distance measures produce several values, as

¹Despite not a conventional network science observable, most important to consider may be the effective distance measure itself, which provides the standard which the model holds itself to throughout its evolution.

²When evaluating effective distance from source to all other nodes with varying source, only the performance of ensembles where ensemble size is comparable or considerably larger than the set of source nodes are considered.

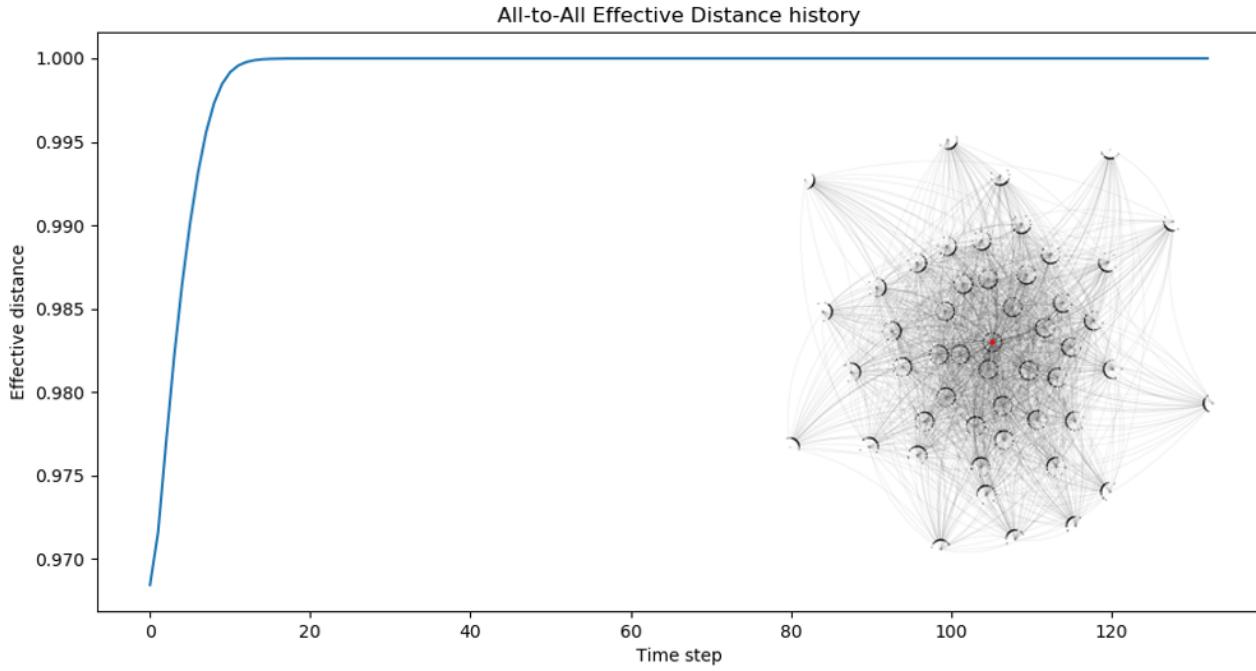


Figure 4.1: Global Effective Distance evolution for constant seeding, uniform random edge initialization, $\phi = 0.4$, $\kappa = 0.15$, showing a long equilibrium state for global effective distance, with the inset network as the final network form.

$\mathfrak{D}_{\text{global}} \in \mathbb{R}^{N \times N}$ and $\mathfrak{D}_{\text{source}} \in \mathbb{R}^N$ and are thus necessarily averaged in order to produce single value, comparable results suitable to grid-search heatmap presentation. Alternatively, rather than considering the difference between initial and final effective distance values, one may consider the mean over all effective distance values through every timestep of the network's evolution. This will naturally skew values to the equilibrium, as over a given run-time there are considerably more near-equilibrium network topologies³ present through the run time than otherwise. Thus averaged network topologies assumed as the edges are continually adjusted yields a more representative picture of the network's performance.

$$\mathfrak{D}_{\text{global}} = \sum_{i,j} \mathfrak{D}_{i,j}^{\text{RW}}(t_f) - \sum_{i,j} \mathfrak{D}_{i,j}^{\text{RW}}(t_i) \quad \& \quad \mathfrak{D}_{\text{local}} = \sum_i \mathfrak{D}_{i,j=s_f}^{\text{RW}}(t_f) - \sum_i \mathfrak{D}_{i,j=s_0}^{\text{RW}}(t_0)$$

Where s_0, s_f are the final and initial source nodes respectively.

The initial network topology has the largest impact on both the development and performance of the network, as though the process of network evolution may be identical, should one grid-search be initialized with a uniform random edge initialization, as is considered the base case and reflected in figure 5.8, and another grid-search is initialized with a sparse edge initialization, not only is the adaptation itself considerably limited, as the Source to Structure model does not allow for edge rewiring⁴, only reweighting existing edges, but the network may only adapt *away* from an already nearly ideal system. This evolution is nonetheless remarkable, as it suggests that the model will adjust *away* from an optimal solution if so initialized and within the relevant parameter regime.

³The networks are seen to approach equilibrium asymptotically (see figure 4.1) and thus even when their run-time is cut-off upon reaching the equilibrium condition, there will be many more values considered asymptotically close to equilibrium than in the initial phase of rapid network adjustment.

⁴This restriction on edge rewiring is effectively relaxed in the case of a complete graph, as all edges which *could* be are already present.

4.1.2 Degree Distributions

An essential feature of the Source to Structure model is the conservation of attention (in-degree) or broadcasting ability (out-degree)⁵, as well as the conservation of the initial edge structure, as only edge weights may be adjusted. Thus though some edges may be nearly normalized out of relevance, no new edges can emerge that were not present in the edge initialization. These features, combined with the weighted, directed nature of the networks in consideration ensure that any degree-based analysis or observable is interpreted differently than its unweighted, undirected equivalent. For instance, hubs with high degree in symmetric, undirected networks imply a reciprocated relation between the connected nodes, with routing possibilities through this central hub extending both to and from the node; e.g. in the case of a single central node to which every node is connected, every node in the network is at most two steps to any other node. [25, 30] Naturally, in the case of a directed node with connections extending *from* itself *to* all other nodes, this central node may itself be one hop from every other node, however all other nodes cannot route through the central node, and thus the global distance between any two nodes is considerably greater than in the undirected case. These observations may appear rather trivial, but are nonetheless of relevance given both their importance in developing and analyzing the model's meta-metrics, and more readily interpreting potentially deceiving graphical depictions of the network.

In order to refine the central set of observables, many degree-based observables were implemented, including the non-normalized degree distribution⁶, edge degree distributions yielding the number of edges with a particular weight, average neighbor degree, and the corresponding variance for neighbor and node degree. [54] Given that either in or out degree is always conserved, and thus so is the total degree, the average degree is likewise constant [25].

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in} = \langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{out} = \frac{L}{N} \quad (4.1)$$

Where $N = |\mathcal{N}|$, k^{in} and k^{out} are the in and out degree, respectively, and $L = \sum_{i=1}^N k_i^{in} = \sum_{i=1}^N k_i^{out} = N$ is their respective totals, which for normalized and conserved degree, will naturally equal the number of nodes present in the system. The degree distribution is given as $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$, where p_k is the normalized degree distribution which gives the probability that a given node has degree k . The average neighbor *in*-degree⁷ is simply the average of the sum of the neighbor *in*degree, given as

$$k_{nn,i}^{in} = \frac{1}{s_i} \sum_{j \in nbr(i)} \mathbf{A}_{ij} k_j^{in} \quad (4.2)$$

where the weighted degree s_i of node i is always one as shown earlier, and $nbr(i)$ is the neighborhood nodes of node i . The variance of the degree distribution is a convenient measure which helps describe the polarization of the degree distribution and the corresponding graph, with high variance of the degree distribution reflecting a wider spread of the edge values, and thus an increasingly specialized network.

4.1.3 Cluster Coefficient

The cluster coefficient is a measure in the range $(0, 1)$ of the connectedness of a given node's neighbors, where a value of zero means that neighbors of the node in question possess no direct connections between each other, and while a clustering value of 1 means that the node's neighbors are entirely connected, i.e. there is a connection between every neighbor to every other neighbor. [25]

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad \langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (4.3)$$

⁵While *in* degree is conserved for all results shown here, a discussion of the effect of switching all binaries (including whether in or out degree is conserved) may be found in the appendix, 7.4.1, and results in the associated git repository [54]

⁶The non-normalized summed degree distribution considers the sum-total non-conserved-degree of each node, $N_k = p_k \cdot N$

⁷In implementation, this is borrowed directly from the NetworkX python library.

where L_i represents the number of links between the k_i neighbors of node i . The weighted version of the clustering coefficient, first proposed in [61], and updated in [62] has subsequently been more efficiently implemented⁸ and is given by

$$\tilde{C}_{i,H} = \frac{\mathbf{A}_{ii}^3}{(\mathbf{A}\mathbf{A}_{\max}\mathbf{A})_{ii}} \quad (4.4)$$

where \mathbf{A} is simply the weighted adjacency matrix and \mathbf{A}_{\max} is a matrix where every element is given as $\max(\mathbf{A})$. Due in part to its greater computational demands, averaged clustering coefficient histories are considered only for more critical parameter regimes in order to give an indication of connectedness beyond the indeterminate visual estimation of network graphs or the unspecific estimate of degree distribution variance.

4.1.4 Shortest Paths

Evaluated in the context of network structure classification parameters, the shortest paths between any two nodes, taking into account edge weight⁹, may be given via Dijkstra's algorithm [53]¹⁰ in $\Theta(|E| + |V|\log|V|)$ time, which thus limits use to networks or parameter combinations of particular interest, especially when considering averages of all to all shortest paths through the entire run time. Nonetheless, this parameter is incorporated into routing efficiency E_{rout} , which gives a more comparable metric based on the shortest paths, and is detailed in section 4.3. [30]

4.2 Classifying Network Structure

As the objective of the Source to Structure model is to test the natural adaptation of a network without rewiring edges¹¹ and the corresponding efficiency of the resultant network structures, it is useful to consider the existent research into both the formation of hierarchy through evolving models and the performance of those hierarchies. [31, 30] While there are myriad methods of comparing the efficacy of diverse network structures, not least simply via variations in the observables introduced in section 4.1, Corominas-Murtra et al. in [30] define a three dimensional *morphospace* with axes of treeness, feedforwardness, and orderability on which they plot all possible network structures¹², and Goñi et al. [31] define an efficiency morphospace with axes of $\|E_{\text{rout}}\|$ and $\|E_{\text{diff}}\|$, measuring the efficiency of the network according to shortest path routing and diffusion, respectively. Shortest path routing naturally assumes global knowledge of the network, while a diffusion process is inherently a local mechanism of information transfer, requiring nodes only to pass information onto neighboring nodes. These two methods of information transfer through a network thus represent different contextual regimes, wherein if the cost of information is low, routing along the shortest path is likely

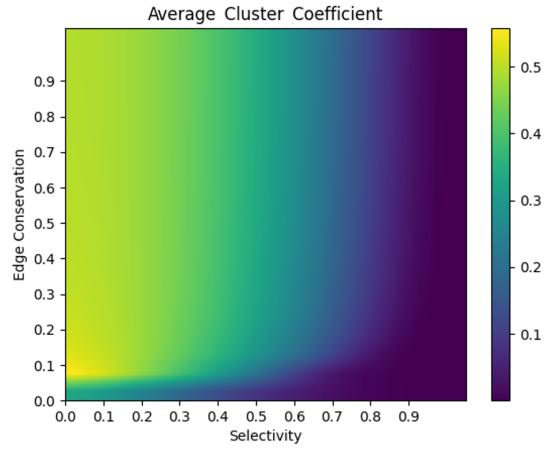


Figure 4.2: Averaged null model cluster coefficients, bilinearly interpolated across $\phi, \kappa \in (0, 1)$, intervals of 0.05 for a connected graph.

⁸The weighted clustering coefficient used is that implemented by the NetworkX python library

⁹It is important to note that path efficiency is considered to *penalize* edge weight, thus edge weights must be inverted in some form before given to the algorithm, e.g. $\mathbf{1} - \mathbf{A}$

¹⁰Initially implemented in $\Theta((|V| + |E|)\log|V|)$ time, later improved by [55] to the stated $\Theta(|E| + |V|\log|V|)$ time

¹¹Here rewiring is understood as the complete dissolution and reformation of some fraction of a network's edges, and not the adjustment of their weights, as is consistent with the adaptation seen in the Source to Structure model.

¹²While plotting a wide range of empirical networks as well as simulated networks, Corominas-Murtra et al. find a wide range of their morphospace absent, and thus discover some network structures which though possible, are never seen *in the wild*. Specifically, they identify networks with a strong hierarchies of cycles as "...belong[ing] to the domain of the possible, but not the actual."

preferred, while if the cost of maintaining or forming connections is high, then a diffusion process is likely preferred, though this does not consider the cost of sending information, which likewise correlates with a preference for optimizing path length. These general, quasi-economic considerations may translate to notions of parallel information transfer/processing, whereby operations performed serially are best performed with global knowledge, and optimal shortest paths, while parallel operations are best performed in regimes of optimal diffusive efficiency. The morphospace of network hierarchy considered in [30] (seen in the left of figure 4.3) classifies all network structures and thus allows for ready comparison of network hierarchy without resort to the monodimensional analysis of individual, yet coupled network science metrics such as global/local efficiency [63], effective distances [52], flow hierarchy [64], rich club coefficients [65], or centrality measures¹³ [66, 67, 68, 69].

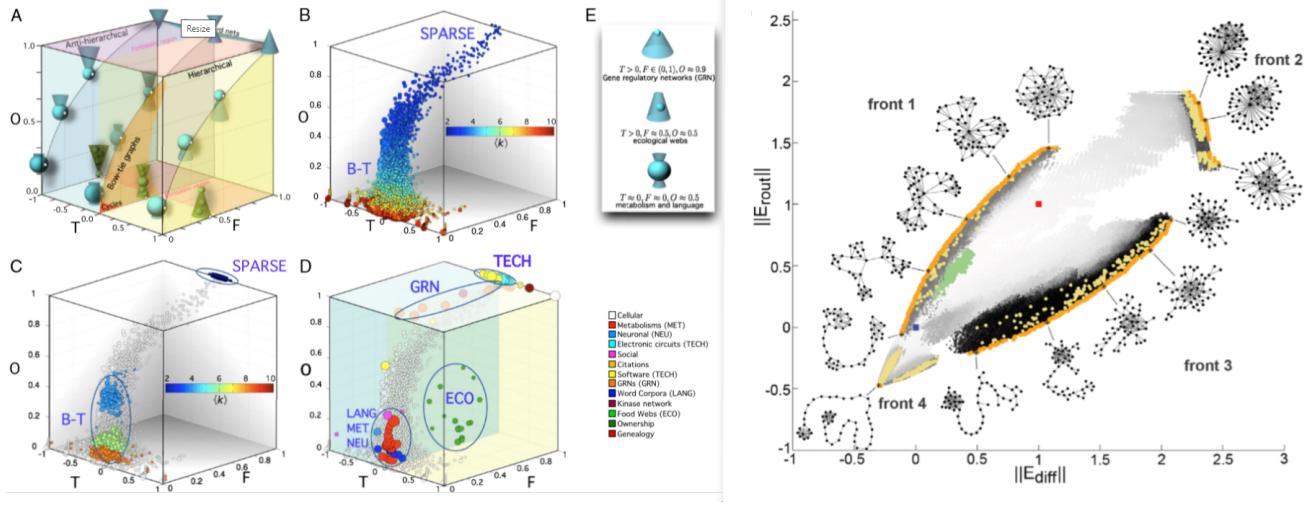


Figure 4.3: Left: Hierarchy morphospace from [30] with axes of Treeness, Orderability, and Feedforwardness, with diagrams displaying intuitive hierarchical depictions in the upper left (A) and real-world networks plotted on the other slides. Right: Efficiency morphospace from [31] depicting efficiency in routing information through shortest paths or diffusion, and depictions of the corresponding network structures. The progression of darkening points from the center into orange *fronts* reflect the exploration of the efficiency space via a simple random-alteration based network evolution.

As both morphospaces proposed in [31, 30] allow for clear classification in terms of other known networks, as well as offer initial theories explaining their behavior and corresponding propensity for natural or technological occurrence, we will classify the results of the Source to Structure model making use of both of their metrics, as well as the intermediate, more well known network science observables when insightful. To this end, we introduce the axes of hierarchical categorization used in [30]: treeness, feedforwardness and orderability. The following is a synopsis of the theoretical description of these observables as introduced in [30], with particular use made of its supplementary material, found [here](#).

Orderability

The orderability of a network describes the interconnectedness of intermediate clusters/communities within the hierarchy. More formally, this is given as the fraction of the entire network which are *not* part of a cycle, or

$$O(\mathcal{G}) = \frac{|\{n_i \in \mathcal{N}_C \cap \mathcal{N}\}|}{|\mathcal{N}|} \quad (4.5)$$

¹³Citations are to betweenness centrality, percolation centrality, eigenvector centrality, and closeness centrality, respectively.

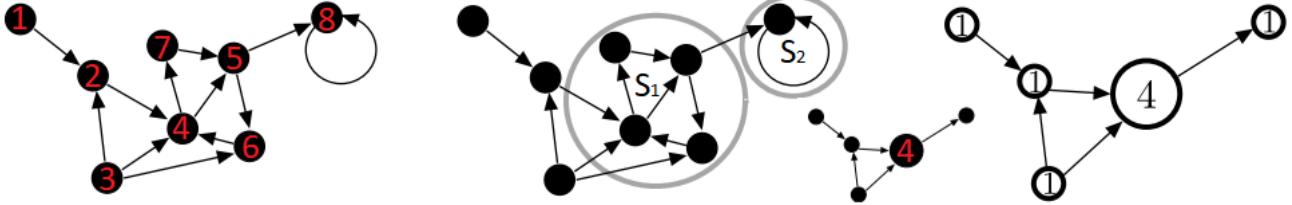


Figure 4.4: Illustration of graph condensation $\mathcal{G} \rightarrow \mathcal{G}_C$, via consolidation of the *strongly connected components* S_1 and S_2 , into a directed acyclic graph preserving the original graph's *directedness*. When condensing the graph, the condensed nodes are themselves given values α_i equal to the sum of the nodes which constituted the compressed cycle. Modified from the *Supporting Information* of [30], where it is featured as figure 3.

where V, V_C are the vertices of the initial, and condensed version of the graph \mathcal{G} respectively. In order to compute the coordinates of hierarchy, a directed graph is first condensed, wherein all cycles are condensed into a single node with a weight equal to the sum of the individual node weights of the cycle. A *cycle* is any directed subgraph whose edges and vertices are components of a walk with equal initial and final points, with the set of all a graph's cycles given as $\mathcal{C}(\mathcal{G}) = \{C_1, \dots, C_k\}$. If a cycle C_k is not a subset of any other cycles,

$$\#C_i \in \mathcal{C}(\mathcal{G}) : C_k \subset C_i \quad (4.6)$$

then it is deemed a *Strongly Connected Component*, and the set of all strongly connected components is given as $\Sigma(\mathcal{G}) = \{S_1, \dots, S_l\}$. The elimination of all cycles (via the concatenation of all strongly connected components, $\Sigma(\mathcal{G})$) in a graph yields a *Directed Acyclic Graph*, or DAG, without cycles but otherwise preserving the hierarchy of the original graph \mathcal{G} , which "...can represent a more or less entangled structure of interconnected causal processes." [30], with the corresponding interconnectivity being represented by the presence of intermediary cycles within the hierarchy. This is illustrated in figure 4.4, where there are two strongly connected components, consisting of

$$\begin{aligned} \mathcal{G}_{S_1} &= \left(\mathcal{N} = \{4, 7, 5, 6\}, \mathcal{E} = \{(4, 7), (7, 5), (5, 6), (6, 4)\} \right) \\ \mathcal{G}_{S_2} &= \left(\mathcal{N} = \{8\}, \mathcal{E} = \{(8, 8)\} \right) \end{aligned}$$

which are subsumed into single nodes via directed acyclic condensation. The relevant adjacency matrices, before and after condensation, are given in equation 4.7. Note that there is never any ambiguity in information flow after condensation, as all cycles may pass information through their cycle, and thus even without any edges apart from those extending to and from the cycle, information seeded into the cycle is assured of making it both to all nodes which are part of the cycle and outside the cycle by whichever node(s) possess edges directed outside the cycle. Within the strongly connected component S_1 there is another cycle,

$$C_1 \subset S_1 = \left(\mathcal{N} = \{4, 5, 6\}, \mathcal{E} = \{(4, 5), (5, 6), (6, 4)\} \right)$$

which illustrates how by considering only the strongly connected components $S_k \in \Sigma(\mathcal{G})$ the subcycles are likewise accounted for within the notion of the hierarchy coordinates here constructed.

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\text{condensation}} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.7)$$

The consolidation of the directed cyclic graph \mathcal{G} into the acyclic \mathcal{G}_C likewise reduces the number of nodes, with cycles becoming individual nodes, a shift which is reflected in the reduced dimension of the adjacency matrices, as seen in equation 4.7.

The orderability $O(\mathcal{G})$ is then simply given as the number of nodes common to both \mathcal{G} and \mathcal{G}_C , i.e. those node which were not affected in any way by the consolidation, and the total number of nodes in the original graph, \mathcal{G} . Making use of the network shown in figure 4.4 we thus obtain $|\mathcal{N}| = 8$, $|\mathcal{N}_C| = 5$, $|\{n_i \in \mathcal{N}_C \cap \mathcal{N}\}| = 3$

$$O(\mathcal{G}) = \frac{|\{n_i \in \mathcal{N}_C \cap \mathcal{N}\}|}{|\mathcal{N}|} = \frac{3}{8} \quad (4.8)$$

Orderability ranges from 0 to 1, with a value of 0 denoting an entirely cyclic network and a value of 1 denoting a perfect line or pyramid, with no communication between those receiving the information and those sending it. Orderability is the most conceptually and computationally simple of the coordinates of network hierarchy, and the condensation of the original graph into the directed acyclic version thereof is a step used in the calculation of both Feedforwardness and Treeness.

Feedforwardness

Feedforwardness is a measure of the network's efficiency in passing information through the entire hierarchy; the fraction of directly hierarchical to indirectly, cyclical connections. Feedforwardness considers paths which extends in the direction of the network's hierarchy, explicitly defined as starting from the set of maximal nodes (i.e. those without in-degree) M to the set of minimal nodes, μ , or those without out-degree. Formally, they are given as

$$M = \{n_i \in \mathcal{N} : k_{in}(n_i) = 0\} \quad (4.9)$$

$$\mu = \{n_i \in \mathcal{N} : k_{out}(n_i) = 0\} \quad (4.10)$$

where M/μ thus being the set of maximal/minimal nodes of \mathcal{G} respectively. We may then consider the set of all paths π_i which extend from M to μ as $\Pi_{M,\mu}(\mathcal{G})$, with a one such path ending in node $n_i \in \mu$ given as

$$\Pi_{M\mu}(n_i) \subseteq \Pi_{M\mu}(\mathcal{G}) \quad (4.11)$$

and the longest such path extending from maximal to minimal node we will denote as

$$L(\mathcal{G}) = \max \left\{ k : \left(\exists n_i, n_j \in \mathcal{N} : \left(\mathbf{A}^k(\mathcal{G}) \right)_{ij} \neq 0 \right) \right\} \quad (4.12)$$

Recalling that after condensing the network as described in section 4.2, every node in the condensed network is assigned as value α_i equal to the sum of the nodes which collectively constituted the compressed

cycle it represents in the directed acyclic version, we may directly write feedforwardness for a single path $\pi_k \in \Pi_{M\mu}$ as

$$F(\pi_k) \equiv \frac{|n(\pi_k)|}{\sum_{n_i \in n(\pi_k)} \alpha_i} \quad (4.13)$$

As an example of the feedforwardness calculation for a single path, if we examine a single path of the directed acyclic graph from figure 4.4 as highlighted in figure 4.5, we find that $F(\pi_j) = \frac{4}{7}$. In order to evaluate the feedforwardness of an entire graph, we first consider the sum of the feedforwardness measures through all maximal to minimal node paths of the condensed graph, $\pi_{ik} \in \Pi_{M\mu}(\mathcal{G}_C)$, giving

$$g(\mathcal{G}) = \sum_{\pi_i \in \Pi_{M\mu}(\mathcal{G}_C)} F(\pi_i) \quad (4.14)$$

and then proceed to 'peel' back layers of the graph through successive elimination of maximal node layers via a leaf-removal algorithm $LR_{f/b}$ and recursively apply the above feedforwardness summation through all the resultant sub-graphs. The leaf-removal algorithm acts as follows; for a given directed acyclic graph \mathcal{G} , all nodes with $k_{in} = 0$ are removed as the first 'leaf', and subsequent iterations sequentially remove the new $k_{in} = 0$ node set. The backwards version of the leaf removal algorithm works identically, though starting and proceeding with $k_{out} = 0$ node sets. Details on the leaf-removal algorithm may be found in [70] [71]. As each removed layer reflects a hierarchical tier, the leaf-removal algorithm will thus naturally produce $L(\mathcal{G}) + 1$ distinct subgraphs $g(\mathcal{G}_1), \dots, g(\mathcal{G}_{L(\mathcal{G})-1})$ composed of all remaining layers. The final feedforwardness coordinate is then given as the average over all subgraphs \mathcal{G}_k :

$$F(\mathcal{G}) = \frac{g(\mathcal{G}_C) + \sum_{k < L(\mathcal{G}_C)} g(\mathcal{G}_C)}{|\Pi_{M\mu}(\mathcal{G}_C)| + \sum_{k < L(\mathcal{G}_C)} |\Pi_{M\mu}(\mathcal{G}_k)|} \quad (4.15)$$

where we note that the summation $\sum_{k < L(\mathcal{G}_C)}$ needn't include $L(\mathcal{G})$ as $\mathcal{G}_{L(\mathcal{G})}$ is by definition pathless, and any single-node graph has feedforwardness 0,

$$\forall \mathcal{G} : |\mathcal{N}| = 1 \implies F(\mathcal{G}) = 0$$

Loosely Feedforwardness represents the length of the paths through the network as a proportion of its cycles, something perhaps best seen, along with the examples of network topology for all hierarchy coordinates, in figure 4.6.

Treeness

Treeness describes "...how pyramidal is the structure and how unambiguous is its chain of command." [29], assuming a value $-1 < T < 1$ which characterizes hierarchy as $T > 0$ and antihierarchy as $T < 0$. Treeness is both the most descriptive and hardest to compute coordinate of hierarchy, requiring calculation of the entire condensed graph's *path entropy*, a measure of the information required to traverse a single path, or in other words, probability not to follow any other path. The number of paths through a network, especially if not otherwise constrained, explodes combinatorially with the size of the network, making the computation of the path entropy (equation 4.21) infeasible for even modestly sized networks. Corominas et al. having already considered this difficulty, have developed a closed exact form of path entropy. [29, 30]

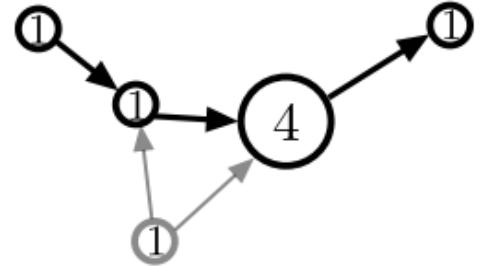


Figure 4.5: The *orderable fraction* of a given path gives the path's *Feedforwardness*, or the number of nodes of a given path $\pi_j \in \Pi_{M\mu}(\mathcal{G}_C)$ divided by their node-weighted sum. Borrowed from the *Supporting Information* of [30], where it is featured as figure 8.

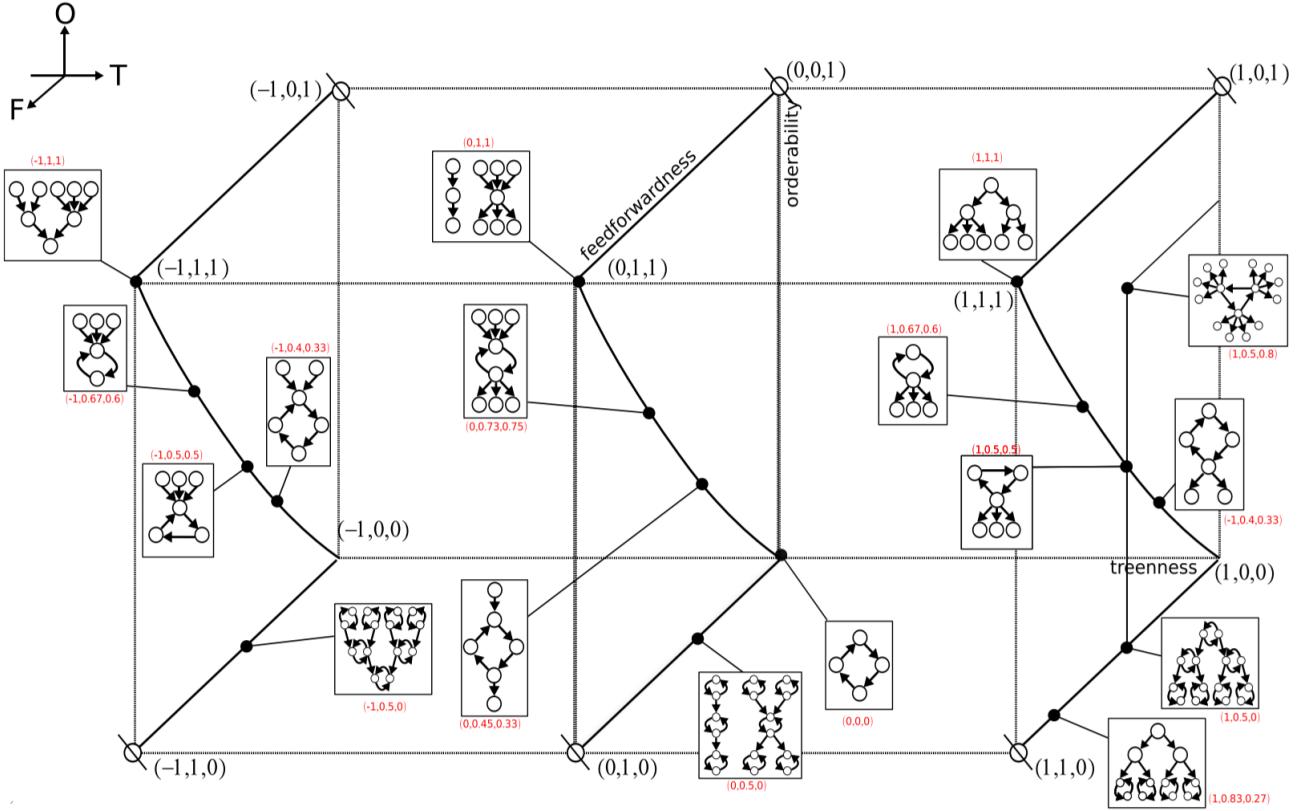


Figure 4.6: Morphospace schematic, with exemplary topologies for distinct points within the morphospace defined by the coordinates (treeness, feedforwardness, orderability). Note that there are regions which are undefined based on the coordinate definitions, as there is no possible topology which could satisfy that parameter combination, e.g. (1, 0, 1). Furthermore, there are regions which though possible to construct topologies satisfying their definition, do not appear in any known real networks. Borrowed from the *Supporting Information* of [30], where it is featured as figure 13.

Let $\mathbf{A}_{ij}(\mathcal{G}_C)$ be the (unweighted) adjacency matrix of the condensed graph \mathcal{G}_C , letting us define the matrix $|V_C \setminus \mathcal{W}_C| \times |V_C \setminus \mathcal{W}_C| \mathbf{B}(\mathcal{G})$:

$$\mathbf{B}(\mathcal{G})_{ij} = \begin{cases} \forall n_i, n_j : A_{ij} = 1 & \frac{A_{ij}(\mathcal{G}_C)}{k_{in}(n_j)} \\ \forall n_i, n_j : A_{ij} = 0 & 0 \end{cases} \quad (4.16)$$

as this ensures any indeterminacy where $k_{in}(n_j) = 0$ is accounted for. The construction of $\mathbf{B}(\mathcal{G})$ then yields the probability of *crossing a node* n_j when following a path from n_i such that, as akin to when considering the row-normalized weighted adjacency matrices as Markov process transition matrices (see section 2.1):

$$\mathcal{P}(n_j \leftarrow n_i) = \sum_{1 \leq k \leq L(\mathcal{G}_C)} \left(\left[\mathbf{B}^T \right]^k (\mathcal{G}) \right)_{ij} \quad (4.17)$$

By considering the uncertainty in a given maximal to minimal path, we find the probability that a given path was chosen of all possible options, while assuming that equal length paths are all equally likely to be chosen. The average (backwards) uncertainty is then

$$H_b(\mathcal{G}_C) = \frac{1}{|\mu|} \sum_{n_i \in \mu} \sum_{n_k \in V_C \setminus M} \mathcal{P}(n_i \leftarrow n_k) \cdot \log(k_{in}(n_k)) \quad (4.18)$$

In order to calculate the *forward* entropy, we simply reverse the definitions of the above $\mathbf{B}(\mathcal{G})$, thus obtaining:

$$\mathbf{B}'(\mathcal{G})_{ij} = \begin{cases} \forall n_i, n_j : A_{ij} = 1 & \frac{A_{ij}(\mathcal{G}_C)}{k_{out}(n_j)} \\ \forall n_i, n_j : A_{ij} = 0 & 0 \end{cases}$$

$$\mathcal{P}(n_i \rightarrow n_j) = \sum_{1 \leq k \leq L(\mathcal{G}_C)} \left([\mathbf{B}']^k (\mathcal{G}) \right)_{ij} \quad (4.19)$$

$$H_f(\mathcal{G}_C) = \frac{1}{|M|} \sum_{n_i \in M} \sum_{n_k \in V_C \setminus \mu} \mathcal{P}(n_i \rightarrow n_k) \cdot \log(k_{out}(n_k)) \quad (4.20)$$

Formally, Treeness is given by computing the path entropy for every path extending from maximal to minimal nodes, $\pi_k \in \Pi_{M\mu}(\mathcal{G}_C)$, equivalently considered as the information required to follow the path, explicitly given by:

$$h_f(v_i) = - \sum_{\pi_k \in \Pi_{M\mu}} \mathcal{P}(\pi_k | v_i) \log \mathcal{P}(\pi_k | v_i) \quad (4.21)$$

where $\mathcal{P}(\pi_k | v_i)$ is the probability that the path π_k (starting at node v_i) is followed. After averaging this *forward entropy* of the network \mathcal{G}_C into $H_f(\mathcal{G}_C)$, one then computes the *backwards entropy* $H_b(\mathcal{G}_C)$ via considering paths Π_{μ_M} from the minimal nodes μ to the maximal nodes M , and then normalizes the difference to obtain the network's proto-treeness:

$$f(\mathcal{G}) = \frac{H_f(\mathcal{G}_C) - H_b(\mathcal{G}_C)}{\max\{H_f(\mathcal{G}_C), H_b(\mathcal{G}_C)\}} \quad \text{where } H_f(\mathcal{G}_C) = \frac{1}{|M|} \sum_{v_i \in M} h_f(v_i) \quad (4.22)$$

while the *treeness* coordinate is given by averaging this proto-treeness measure over the full set \mathcal{W}_G of sub-graphs of \mathcal{G}_C obtained through recursive application of leaf-removal algorithms, such that our final treeness value for the graph \mathcal{G} is given as

$$T(\mathcal{G}) = \frac{1}{|\mathcal{W}_G|} \sum_{\mathcal{G}_i \in \mathcal{W}(\mathcal{G})} f(\mathcal{G}_i)$$

$$= \langle f \rangle_{\mathcal{W}_G} \quad (4.23)$$

To give some intuition to this notion of *treeness* consider that in a linear chain, $H_f(\mathcal{G}_C) = H_b(\mathcal{G}_C) = 0$ as there is neither a spreading of information from a single source, as would be hierarchical, $H_f(\mathcal{G}_C) > H_b(\mathcal{G}_C)$, nor consolidation of information from various sources into a smaller number of nodes, which would be anti-hierarchical, $H_f(\mathcal{G}_C) < H_b(\mathcal{G}_C)$. Corominas-Murtra et al. generously illustrate the different network hierarchies that the *treeness* coordinate is intended to describe in a variety of figures to be found within their paper's appendix, and we copy one such illustration to the in figure 4.7.

4.3 Global vs Diffusive Communication Efficiency

Routing and diffusion efficiency were introduced in [72] and used by Gōni et al. [31] in an examination of network communication efficiency, which classifies networks' efficiency in passing information between any two randomly chosen members of the network through either the shortest possible path or diffusion,

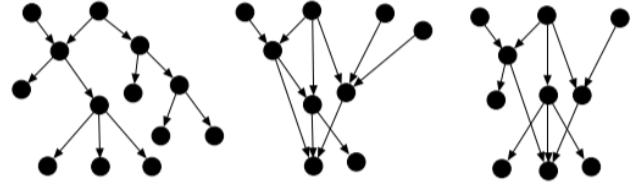


Figure 4.7: From left to right, the networks appear hierarchical $H_f - H_b > 0$, then anti-hierarchical $H_f - H_b < 0$, then neither, as $H_f - H_b \approx 0$. In the center figure, there are clearly more Borrowed from the *Supporting Information* of [30], where it is featured as figure 11.

proto-treeness:

$$f(\mathcal{G}) = \frac{H_f(\mathcal{G}_C) - H_b(\mathcal{G}_C)}{\max\{H_f(\mathcal{G}_C), H_b(\mathcal{G}_C)\}} \quad \text{where } H_f(\mathcal{G}_C) = \frac{1}{|M|} \sum_{v_i \in M} h_f(v_i) \quad (4.22)$$

$$T(\mathcal{G}) = \frac{1}{|\mathcal{W}_G|} \sum_{\mathcal{G}_i \in \mathcal{W}(\mathcal{G})} f(\mathcal{G}_i)$$

$$= \langle f \rangle_{\mathcal{W}_G} \quad (4.23)$$

To give some intuition to this notion of *treeness* consider that in a linear chain, $H_f(\mathcal{G}_C) = H_b(\mathcal{G}_C) = 0$ as there is neither a spreading of information from a single source, as would be hierarchical, $H_f(\mathcal{G}_C) > H_b(\mathcal{G}_C)$, nor consolidation of information from various sources into a smaller number of nodes, which would be anti-hierarchical, $H_f(\mathcal{G}_C) < H_b(\mathcal{G}_C)$. Corominas-Murtra et al. generously illustrate the different network hierarchies that the *treeness* coordinate is intended to describe in a variety of figures to be found within their paper's appendix, and we copy one such illustration to the in figure 4.7.

respectively. Routing, or navigation¹⁴ efficiency measures the ease of communication within the network with only the shortest path between any two nodes used, which implies both global knowledge of the network (so that nodes may identify the shortest paths beforehand) as well as information being passed in parallel [72]¹⁵. Routing efficiency is given simply as the weighted sum of inverted shortest paths,

$$E_{\text{Routing}} = \frac{\sum_i \sum_j 1/\phi_{ij}}{n(n-1)}, j \neq i \quad (4.24)$$

where $\Phi = [\phi_{ij}]$ is the set of shortest paths between all node pairs. As diffusive efficiency is designed to consider the average of all possible paths and their corresponding weighted probability, we may make use of the RWED as introduced in section 2.3, as that is precisely what it considers:

$$E_{\text{Diffusion}} = \frac{\sum_i \sum_j 1/t_{ij}}{n(n-1)} \approx \frac{\sum_i \sum_j \mathfrak{D}_{ij}^{\text{RW}}}{n(n-1)}, j \neq i \quad (4.25)$$

where t_{ij} is the mean first passage time whose inverse roughly equates to the RWED, as $t_{ij} = \frac{\mathbf{Z}_{jj} - \mathbf{Z}_{ij}}{\omega_j}$, $i \neq j$. Both routing and diffusive efficiency measures should be normalized to the scale of the network, according to the average efficiencies of a lattice and randomly connected network of the same size:

$$\|E_{\text{rout}}\| = \frac{R_{\text{rout}} - \langle E_{\text{rout}}^{\text{latt}} \rangle}{\langle E_{\text{rout}}^{\text{rand}} \rangle - \langle E_{\text{rout}}^{\text{latt}} \rangle} \quad \|E_{\text{diff}}\| = \frac{R_{\text{diff}} - \langle E_{\text{diff}}^{\text{latt}} \rangle}{\langle E_{\text{diff}}^{\text{rand}} \rangle - \langle E_{\text{diff}}^{\text{latt}} \rangle} \quad (4.26)$$

Together, routing and global efficiency metrics offer a metric for determining the all-to-all efficiency of a network, and the corresponding network topologies which develop through network simulation, of particular relevance when considering whether the Source to Structure model leads to more or less optimal network configurations for overall communication. As many model configurations allow for a mere handful of information sources throughout the simulation, all agents optimizing for proximity to this limited set of information sources would be expected to sacrifice global efficiency, though this may be subverted for intermediary parameter regimes wherein limited connections or high edge conservation could prevent convergence to perfect star-like configurations.

	clique	chain	ring	lattice	star	rich club	bi-modular
<i>n</i>	50	50	50	50	50	50	50
$\langle k \rangle$	49.00	1.96	2.00	6.00	1.96	13.00	24.04
$\langle \text{spl} \rangle$	1.00	17.00	12.75	4.59	1.96	1.98	1.98
E_{rout}	1.00	0.14	0.15	0.34	0.52	0.59	0.67
Π_{spl}	0.02	0.04	0.04	0.03	0.04	0.03	0.02
E_{diff}	0.02	0.01	$<10^{-2}$	0.01	0.03	0.03	0.01
E_{res}	0.03	0.07	0.07	0.05	0.05	0.04	0.03

Figure 4.9: Idealized network topologies and corresponding graph and efficiency metrics; of particular relevance is diffusive and routing efficiencies, E_{diff} and E_{rout} respectively. Borrowed from [31], where it is featured as figure 2.

¹⁴Alternatively *global efficiency*, due to its requirement for global knowledge of the network.

¹⁵In this case, the parallel vs sequential notion of communication considers the difference between evaluating this shortest path efficiency globally and locally, whereby locally disconnected clusters could hamstring a sequential route.

Broadly, the observables considered in the evaluation of the networks produced by the Source to Structure model are summarized below in table 4.1.

<i>Single-Simulation Metrics</i>	<i>Grid-Search Metrics</i>	<i>Grid-Search Heat Maps</i>
Eff. Dist. Differences	Hierarchy Coordinates	Degree Distribution σ
Source Distribution		Average Neighbor Differences
Degree Distribution	Efficiency Coordinates	
End Edge Weights		Effective Distance Differences
Network Evolution gif		(To source to all to all)

Table 4.1: Table of observables, as used for various scopes of investigation.

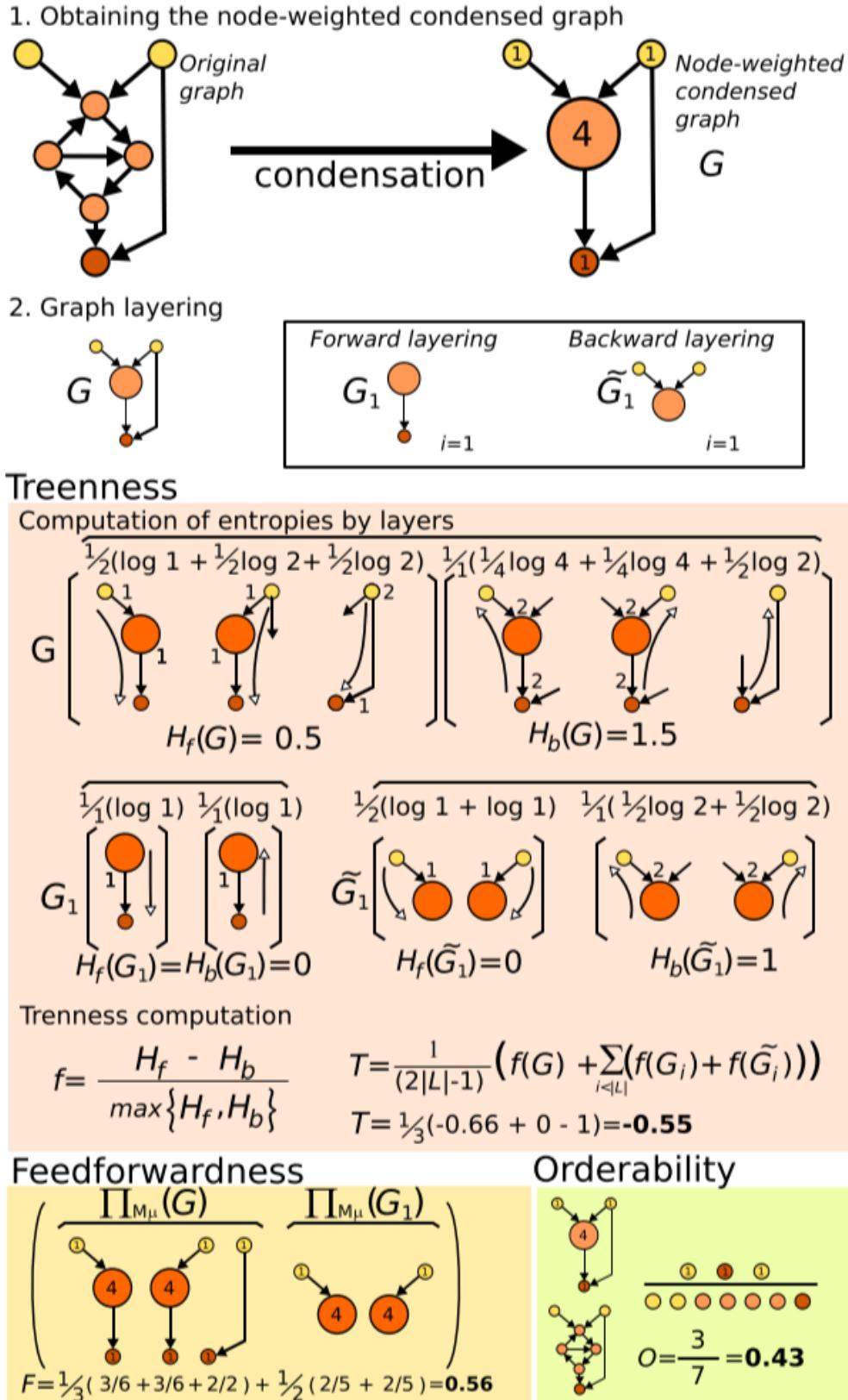


Figure 4.8: Info graphic visually describing the process of graph consolidation, layering and hierarchy coordinate computation. Borrowed from the *Supporting Information* of [30], where it is featured as figure 11.

Results

5.1 Model Configurations

Given that the Source to Structure model is intended to reflect abstracted notions of network adaptation to information dissemination, there are many possible scenarios reflected by various configurations of the model. Broadly, there are five different points wherein the model may be modified to account for the diversity of empirical phenomena; binaries, initial network structure, source selection, system parameters and model mechanics tweaks, and though all variations are briefly summarized here, only the most interesting cases are examined in detail in this work. Binaries refer to fundamental alterations to the model¹, network initialization determines the unchanging set of starting connections, though their weights may change through the simulation, while source seeding refers to distribution of information founts over the course of the simulation. System parameters are explored via grid search, and account for the psychological heuristics or environment of the agents, while model mechanics are optimized for ease of computation and performance independent of the model configuration. While discussed in detail in sections 5.1.1 - 5.1.4, the set of model variations which collectively constitute model configurations are summarized in table 5.1. As binary modifications are not investigated thoroughly in the following results, description thereof may be found in the appendix, section 7.4.1.

<i>Binaries</i>	<i>Edge Initialization</i>	<i>Source Seeding</i>	<i>Run Time</i>	<i>System Parameters</i>
MPED vs RWED	Uniform random Sparse	Random Constant Source	Update Interval Source Reward	Number of nodes
Info Score to Reinforcement Coupling	Scale-Free $(P_k = k^{-\gamma})$	Normal Distribution	δ Parameter	Edge Conservation Coefficient
	Empirical Network	Power Law Distribution		Selectivity
Inverted Eff. Dist. Correlation		Diversity of Connection		

Table 5.1: All variations of the implemented model; the diversity these modifications are crafted to suit the empirical phenomena, and ideally allow for high model fertility (relevance to a wide variety of phenomena) with minimal adjustments.

¹These fundamental shifts include changes in directionality, such as shifting all agent's optimization to be based on effective distance *to* the source, rather than *from* the source, or presuming information propagates inversely, or conserving inward rather than outward connections. As these changes allow for a vast array of different model behavior, only a subset of possible behaviors (those most relevant discovered through other parameters) will be highlighted of behavior discovered through variance of the other parameters.

At first glance, the variation allowed by these various configurations might lead one to suspect that with sufficient tuning of the model's various parameters, one could be able to produce nearly any result, thus rendering the model's predictive capacity and interpretation moot. The plethora of parameter permutations do not however change the core model methodology, as they either allow for more ready comparison to empirical forms, as in the case of varying initializations and seeding mechanisms, or else are designed to account for a variety of agent heuristics, as the parameter grid-searches explore. Though much of the model's diverse behavior is directly accounted for by the quantity of its configurations, the effect of the core of the model's run process remains consistent, and thus further configurations serve better as further testing of this core mechanic rather than an invalidation of the model due to potential excess flexibility. Alternatively, all variations of the present model (especially those not considered in grid-search) are the result of expanding options of what would otherwise be an arbitrary decision; e.g. various network initializations does not so much allow for greater flexibility to influence the outcome towards intended results as it simply expands the set of scenarios the model may reflect.

5.1.1 Model Mechanics

There are certain variables which though essential to the core processes of the Source to Structure model are not analogues to empirical phenomena or proxies for agent behavior, instead serving as minor variations in the model environment. These variables are thus simply optimized to allow for the clarity of other parameter variations, and otherwise left unchanged; both the *source reward* and constant δ were examined both analytically and through exhaustive computational testing and then set to the values which allow for the clearest variation of other parameters while preserving the model's computational integrity, i.e. avoiding floating point errors while accelerating relative rate of network adaptation and preventing disproportional intermediate exponential suppression.

- δ Parameter: Remains largely invariant, but originally refers to transition rates in the SIR model².
- Source Reward: Reward of source proportional to the next-best. (e.g. a source reward of 2 implies that the source is given an effective distance half of the smallest effective distance from itself.)
- Edge reweight coupled to info score: Determines if reinforcement is proportional to the info score.³
- Update Interval: Number of seeding steps before edge update

The " δ -parameter" appears in the derivation of the RWED (equation 2.26) as a conglomeration of SIR parameters, explicitly as

$$\delta = \ln\left(\frac{\beta - \mu}{\alpha}\right) - \gamma_e$$

and in the final RWED (equation 2.43) as

$$\mathbf{D}^{\text{RW}} = -\ln \left[\mathbf{Z} \cdot \text{Diag} \left(\mathbf{Z}^{-1} \right) \right] \quad \text{where } \mathbf{Z} = \left[\mathbf{1} - \mathbf{T} e^{-\delta} \right]^{-1} \quad (5.1)$$

thus δ simply suppresses the values of the row-normalized adjacency matrix (which is equivalent to the Markov transition matrix, \mathbf{T}) for values of $\delta > 1$. This suppression has little overall effect as it's applied evenly (see git repository, for a detailed review of the effect of varying δ on model results), and thus δ has

²This parameter is inherited from use of Koher et al's [32, 52] effective distance (most notably RWED) methods

³This is especially relevant for selectivity=0 cases, as then the only thing differentiating rewards for nodes is their respective info scores.

been fixed at 10 (which showed the clearest evolution) for all results shown.⁴

The source reward is the divisor used to determine the effective distance of the source itself, which naturally would be 0 otherwise, as the source is 0 distance from itself. For info-scores to be well defined, all effective distances must be non-zero, and so the source must be given a non-zero value which still reflects its position as the closest to the source, and thus the source reward is used as the divisor of the lowest effective distance, insuring that the source retains its position as the lowest effective distance. Variation of the source reward has generally minimal overall effect in the evolution of the resultant network structures, which itself is a non-trivial result as best seen via direct comparison (data compiled in the associated git repository), though some effect may be observed for source rewards less than 2. This generally means that so long as there is at least a 100% factor delineating the source from its nearest neighbors, the model evolution proceeds similarly. For this reason, and to reduce overall complexity, the source reward has been set to a constant 2.6 for all simulations producing examined results seen in section 5.2.

Whether edges are reweighed directly proportional to their info score or are instead rewarded by a constant factor if their info score is above a given threshold relative to the node’s other edges, does not have a great effect on the overall evolution of the network during simulation. It is however a potential psychological distinction, determining the granularity of the agent’s response and whether or not every agent must be able to distinguish the relevance of all of their connections or are simply broadly attempting to better whichever connections perform well. This is most relevant and apparent in the edge case of selectivity $\phi \approx 0$ as then if all agents are rewarded equally (as selectivity is low, all agents are then considered for reinforcement) agents will not be able to adjust their connections, being apparently too indecisive and egalitarian in their reinforcement.

The number of source-diffusion steps taken before agents are allowed to re-evaluate their edges, or *update interval*, allows for a smoothing out of inconsistent source seeding by yielding the summed info-scores which accord the various sources seeded within the update interval. Though this behavior was observed in preliminary tests, the update interval was not adjusted from 1 through the entirety of the presented results, so as not to confound the initial investigation. Nonetheless this update interval has a clear relation to the timescale of the information sourcing and its related dispersion, as longer update intervals indicate a slower relative rate of inter-network communication (via diffusion), which find use in replicating the effects of evolutionary regimes of real world networks.

5.1.2 Edge Initialization

Most real world networks exhibit a characteristic degree distribution resembling one of those abstracted versions well studied in network science, most commonly scale free (i.e. obeying a power-law degree distribution $p_k = k^{-\gamma}$) with average degree $1 < \langle k \rangle < \ln(N)$ [36, 25]. In order to most readily resemble real-world networks as well as provide for unrealistic network structures to test model dynamics and ideal network structure, the Source to Structure model may be initialized with a variety of network structures as listed below. In the context of the Source to Structure model, initializations consist purely of establishing the edges and their associated weights for a set number of nodes. For all initializations the (row) normalization condition holds, and the networks may be optionally ‘undirectified’ by averaging all reciprocated edges.

- Uniform: uniform random edge initialization; creates a fully connected all to all small world network.
- Scale-Free: Edges are initialized with the probability of any given node having k connections given

⁴One may ask why not eliminate δ entirely if serves little purpose and is set constant throughout the model’s exploration, however as δ is derived from SIR dynamics (which are values specific to the virility of a pathogen, rate of agent contacts, and environment generally, by leaving δ as part of the model it is conceivable that with sufficient specificity of all other parameters, one may solve for δ , providing further insight into the environment of the agents. This was however not explored through the course of this thesis, and conceivably was not worth the computational cost, though $\delta = 10$ did accentuate trends which were otherwise more computationally cumbersome to discover.

by the power law $P_k = k^{-\gamma}$, where γ is then the degree exponent.⁵

- Sparse: initializes every node with a set integer number of randomly assigned connections per node with approximately equal weight.⁶
- Real-World: Initialize with (potentially scaled) versions of real-world networks.

Though highly unrealistic, uniform random edge initialization serves as a general base case, as it allows for any network structure to evolve as all nodes are initially connected to all others.⁷ All to all edge initialization (complete graph initialization) permits the development of structures which rely on global knowledge, though the depth of this knowledge is limited by the wider distribution of edge weights⁸ as evidenced by the case of a constant source node, with conducive selectivity and edge conservation a strict star-like structure may be formed, wherein all nodes are directly connected to a constant source. Real-World networks generally obey a power law degree distribution in both number and weight of connections [36], however as a uniform distribution of edges and their weights was found to be nearly globally optimal any local optimization which alters the initial network structure suggests either counter-productive adaptation⁹ or else reflects an environment wherein the source is sufficiently consistent such that local connections are more efficient than global network communication. As the Source to Structure model does not allow for edge rewiring (the creation of new edges), uniform random edge initialization is unique in that all possible structures *may* be formed therefrom, which makes it suitable as the default initialization method for examination of the model mechanics and network dynamics.

Scale free edge initialization provides for the most ready comparison to a wide range of real world networks characterized by their degree exponent γ , though the comparably small networks over which most simulations were run permits only an imperfect resemblance to the power law degree distribution, as in order to maintain a connected network all nodes must retain at least one connection, and the distribution of edge weights themselves are even across a given node¹⁰, as seen in figure 5.1. By default scale-free *out-degree* is examined, though in cases wherein the relevant binaries have been switched, (i.e. edges adapting their out-degree) scale-free in-degree initializations may also be considered. Given the comparative lack of connections provided each node in the case of power law degree distributions, the selectivity parameter has far less effect than in the uniform random case, as nodes may more readily distinguish between their edges by virtue of their comparative strength (see equation 5.2), eliminating the purpose of the selectivity parameter ϕ .

$$\text{average edge weight} = \epsilon = \frac{1}{\langle k_{\text{all-to-all}} \rangle} = \frac{1}{\frac{1}{N}} < \frac{1}{\langle k_{\text{scale-free}} \rangle} \quad (5.2)$$

Given the reduced set of connections overall in comparison to uniform random edge initial network structures, every network must strive for an imperfect optimum while the effects of selectivity and edge conservation are obscured through the relative lack of choices given every node through their reduced set of edges. While the evolution of a small world network may lead to a star-like structure, scale-free networks'

⁵In order to ensure network connectivity and that all nodes are relevant, there is also a minimum number of connections per node, set to one throughout all results shown.

⁶Every connection is of *approximately* equal weight, as in order to preserve the network's connectedness (necessary to the core functioning of the model and its corresponding analysis) if there are many connections per node it may be necessary to allow overlapping weights; though this is more relevant for scale free network initialization.

⁷This effectively relaxes the restriction on edge rewiring, as in this case no *new* edges need be formed when they are all initially present.

⁸As the quality of information transmitted by an edge is proportional its proportion of a node's total edge weight (universally 1, given the normalization condition) uniform random edge distribution gives less detailed global knowledge as number of nodes increases.

⁹Counter-productive adaptation is behavior which leads to worse global and local outcomes, and is thus the purely product of the agent's mechanics, and not an optimization methodology.

¹⁰While conservation of attention is a core element of the model, further investigations may be inclined to allow for varying distributions of total attention allotments to various nodes, so that high degree nodes may not necessarily have lower edge values.

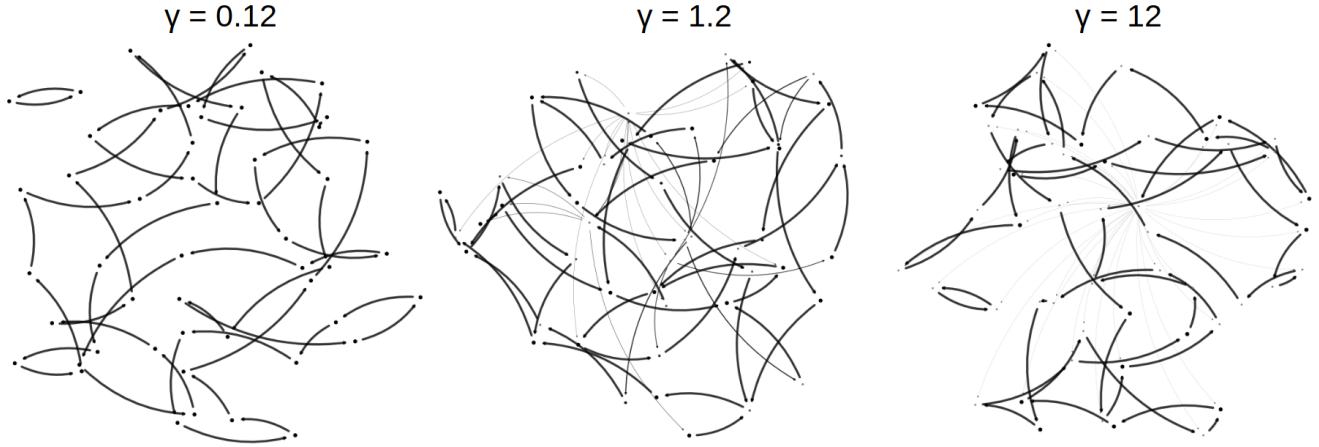


Figure 5.1: Scale free edge initializations, with the degree exponent proceeding from $\gamma = \{0.12, 1.2, 12\}$ from left to right. Note that as the degree exponent increases, there are a smaller fraction of node which possess a larger number of edges, though as all nodes possess the same total edge weight, the weight of each edge is given as $\epsilon = 1/k_i$, with i denoting the node index.

evolution, when given otherwise identical parameters leads to all nodes reinforcing their connection to the highest degree node, which approximates a star as well as possible given the initial network structure. Furthermore, it may well be that such behavior which incidentally optimizes for global communication may perform well for a variety of source seeding scenarios (discussed in section 5.1.3) and thus scale-free networks of diverse parameters converge to the same solutions in an apparent instance of convergent network evolution, while the more general case of uniform random edge initialization allows for optimization away from their default, globally optimal solution to be more effective in a given regime (as in the case of constant seeding). Thus while global optimum are the best a network of fixed structure may achieve, there will be convergence to a global optima, though if networks begin with a nearly globally optimum configuration, they may divergently evolve into more specifically efficient solutions as their structure allows.

By initializing networks with universal, constant degree, sparse edge initialization may be used as a base of comparison to scale-free edge and all-to-all networks while admittedly bearing little immediate resemblance to real-world networks. This disconnect from real world networks is largely due to the perfect homogeneity in contact count across all nodes, though there is evidence that this quantitative limitation [73] is reflected when social networks are considered near enough to their entirety. [57] As the constant degree may be set to the nearest integer of the median/mean degree of a scale-free network, $\langle k_{\text{sparse}} \rangle = k_{\text{sparse}} = \langle k_{\text{scale free}} \rangle$, sparse network structures provide ready comparison to scale-free networks, functioning largely to eliminate the 'central' outlying highest degree nodes which appear in any power law degree distribution. Constant degree networks thus have no 'central' node through which they may begin to establish a star-like structure, nor readily pass information through, thus forcing the development of and reliance upon structures of higher order (multi-step) communication. Given these limitations, and that results suggested that the incomplete *super* nodes¹¹ as seen in scale free networks produced similar behavior to the set-density sparsely initialized networks, later investigation focused on variable density networks via sparse initialization, specifically $k = (0.1, 0.25, 0.5, 0.75, 1) \forall n \in \mathcal{N}$.

Though real world networks with known degree exponents are imitated on smaller scales, no direct empirical network was initialized in these initial simulations.

¹¹i.e. those nodes with the dominate number of connections in a scale-free network.

5.1.3 Source Seeding

The choice of frequency and location of sources throughout the simulation is considered *source seeding*, and can be manipulated to imitate real-world information dispersal; a constant or select few nodes serving as the only sources throughout the simulation may be considered a ready analogue to a professional media outlet regularly serving as the broadcaster of information whose original source lies outside the scope of the network. Alternatively, the same scenario, i.e. a concentrated set of broadcasters, is also analogous to a research institution being responsible for the great majority of papers in a particularly (presumably nascent) academic field. By shifting the source throughout the course of the simulation, the network necessarily adapts if it is still sufficiently flexible to do so; evenly spreading weight across edges provides for globally optimal communication efficiency, which in the context of a rapidly and randomly shifting source, is likely optimal, whereas for a constant source networks may (ideally) connect directly to receive all information directly from the source without 'fear' of loosing information through a shift and necessarily receiving information through a secondary connection. All other seeding mechanisms can thus be seen as intermediaries between these two possibilities, wherein power law and normally distributed source seeding provide for an imperfect consistency to the source position. It is also possible to combine various seeding methods within the same simulation, wherein initially the source may be constant, and after the network has adapted to such consistency, adjust seeding¹² and observe the resultant network adaptation; the more specialized the network structure before the source seeding shift the more difficult to adapt thereafter. In a similar vein, the update interval (how many source-seeding events occur before nodes adapt their edges) likewise has an effect on the specificity of the eventual adaptation; if every node considers many different sources before adjusting their edges, the resulting network structure is likely less specialized.

- Constant Source: Seeds source as the same node for every step.
- Random: Randomly chooses a new source every step.
- Set Number of Shifts: Akin to constant-source, though with a set number of random shifts to unique source nodes throughout the run.
- Normal Distribution: Seeds source via preset normal distribution, with designated kurtosis (σ) and mean.
- Power Law Distribution: Seeds based on scaled power-law, $n \cdot P(x; a) = n \cdot \lfloor ax^{a-1} + 1/2 \rfloor$, $0 \leq x \leq 1$, $a > 0$ with a the power law exponent
- By Diversity of Connection: Seeds proportional to how widely distributed each node's edges are.

Maintaining a single node as the source provides sufficient consistency for the network to adapt to this most specific of cases whose real-world analogue is whenever information generated outside the scope of the network is passed into the network through a single source, as in the case of emergent topics¹³. There is a natural tension between optimization for a specific case, as epitomized by the seeding of a single constant source, global communication efficiency, which naturally is more suited to accommodating shifts in source seeding, as seen in figure 5.2. In network structures without such freedom to specialize however, network evolution accommodating a constant source may not be sufficiently specialized to pass into a globally sub-optimal regime, and instead remain more effective for a variety of source seeding scenarios, as discussed in sections 5.1.2, 5.2.1. Shifting the designated source node randomly at every step prevents

¹²In the most simple case source seeding may be simply adjusted to another node, though switching to another less consistent seeding mechanism forces the network to adapt from a very specifically optimal solution to a more globally optimal solution as the source becomes more global in its distribution.

¹³Emergent topics are likely to be introduced by a single source, before other members of the network adapt and establish their own independent connections to the original external information source, leading to a more power-law like distribution of the source.

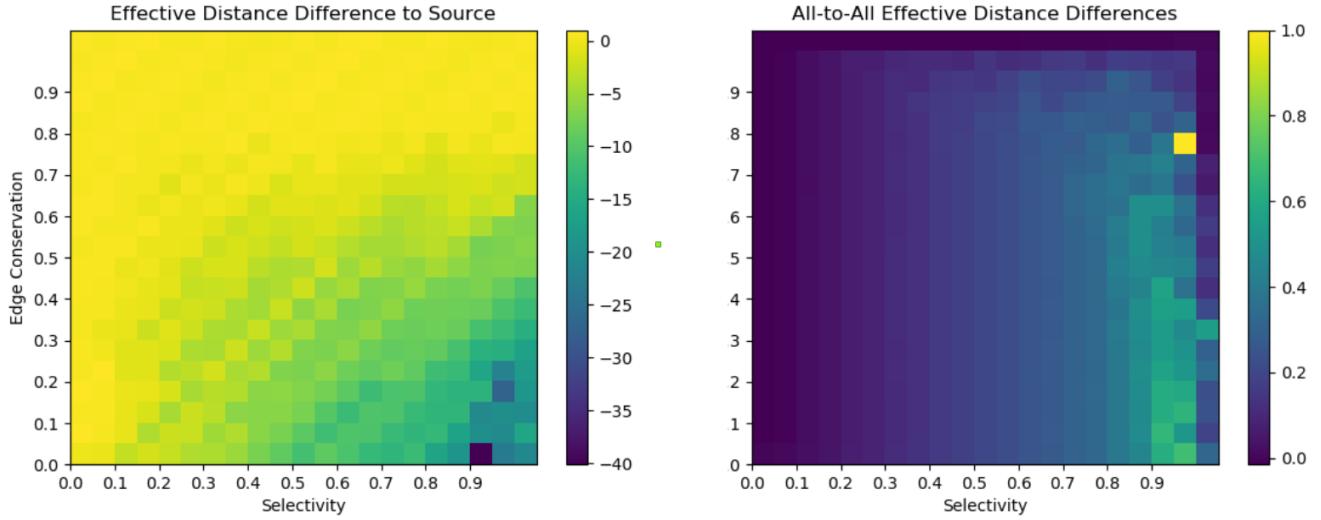


Figure 5.2: Local (left) vs global (right) average effective distance differences compared for constant seeding, uniform random edge initialization, max. 600 runs, $n = 50$, $\phi, \kappa \in \{0, 1\}$, 0.05 intervals. Though not perfectly correlated, it is clear that the more efficient local configurations seen in lower right of the left figure come at the expense of globally optimal solutions. Lower numbers indicate improvement as $ED_{\text{diff}} = ED_{\text{final}} - ED_{\text{initial}}$, and lower (averaged) effective distances indicate more rapid information dissemination.

the network evolution from converging to a specific source-centered form, and thus the resultant structure is likely to closely resemble a more globally optimal solution. The rate of edge adaptation will however determine whether the latest network orientation is configured to whichever node was latest assigned as the source, or else due to the rapid, continuous shifting more slight edge adaptations are canceled out through the course of a given simulation.

Though all other source seeding mechanisms are effective intermediaries between random and constant seeding, the trivial intermediary consists of a set number of shifts over the course of the simulation. By shifting the source node only after the network has already adapted to a consistent source, the rate of network adaptation to a new source from an equilibrium configuration may be observed.¹⁴ When adjusting source seeding via a set number of shifts, unless the number of shifts exceed the number of nodes, the newly assigned source node will be unique (not having been previously assigned as a source node).¹⁵

Normal and power law distributed source seeding both serve as analogues to the most realistic scenarios of information diffusion, wherein a logarithmic subset of the population serves as the source of the greater portion of information generally. [15, 12] Specifically considering digitally leveraged socialization as per microblogging, Bruno Gonçalves et al. [58] show that as the biological and cognitive constraints commonly encapsulated in *Dunbar's number* [73] lead to a quasi-logarithmic response rate as high-degree agents reach their natural attention limits.

Though not investigated in detail, market behavior has also been considered through the lense of threshold models [74, 75] and it is not unreasonable to consider a minority of market agents whose early actions signal market movements which may be observed as 'diffusing' through the market's awareness, as the

¹⁴Note that the rate of network structural adjustment *from equilibrium* may be distinct from adaptation from the 'default' network configuration as given by the network edge initialization.

¹⁵Furthermore, the frequency of source node shifts must be faster than the equilibrium cutoff length, so as to ensure that the simulation continues to run through the course of all intended source shifts and the resulting observable averages are all taken consistently over the course of the full run time for all network configurations.

Source to Structure model *may* effectively represent.¹⁶ By seeding consistently to the same set of nodes

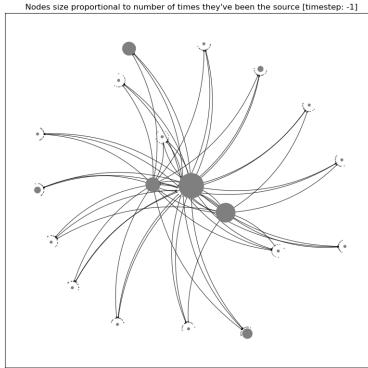


Figure 5.3: Power law source seeding: nodes are sized based on the number of times they have been the source. This graph is shown after 300 runs, $n = 20$, $\phi = 0.8$, $\kappa = 0.2$, $a = 10$ from power law distributed seeding.

of size $\propto \log(n)$, the network is given an opportunity to adjust to normal and power law source seeding, forming structures which are hybrid specializations for locally efficient regimes, as in figure 5.3 which started from the globally optimal uniform edge initialization.

Seeding by diversity of connection was implemented [54], however as it consists of a complex coupling of network structure and re-enforcement, such a source seeding mechanism was deemed beyond the scope of this work, and not thoroughly investigated.

5.1.4 System Parameters

As the number of nodes, selectivity and edge conservation are parameters integral to the evolution of networks though any configuration, grid-searches are run over selectivity and edge conservation coefficients through values extending through their entire range, $\{0, 1\}$ in intervals of 0.05. Though these grid-search parameters reveal the model's behavior through the entire selectivity/edge conservation space, the most clear network evolution is generally observed within higher ranges of selectivity and lower ranges of edge conservation. Grid-searching through the entire parameter space shows the rate of transition into the characteristic behavior confined to the latter ranges of the selectivity-edge conservation space, behavior which is often obscured when considering only diagrams of the resultant graphs after relaxing to their equilibrium form.

- Number of nodes, n
- Edge Conservation Coefficient: κ in info score evaluation via

$$I_{ki} = \left(\frac{n_i}{\sum_j n_j} \right)^{\kappa-1} \cdot \text{edge}_{ki}^{\kappa} \quad (5.3)$$

- Selectivity: ϕ lowest fraction of edges dismissed when considering edge rewards

Though the number of nodes has a marked effect on all parameters, preliminary grid searches over node number as well as selectivity and edge conservation coefficients suggest that for the observables discussed in section 4.1 scale with the log of the number of nodes, as seen in figure All simulations are performed with $n = 60$ unless otherwise specified to allow for sufficient nodes for relevant results, while

¹⁶Further evidence for this broad analogy may be given by the corresponding power laws of market order volumes, returns [76] which vaguely resemble the distribution of information as it diffuses from the source.

preserving possible visualizations and ensuring reasonable computational requirements. Certain features (such as hierarchy coordinates) are best seen through higher node numbers and thus for relevant visualizations isolated runs of configurations of interest are rerun with higher node counts, preventing the need for grid-searches through nodes throughout.

5.4.

Edge conservation κ determines whether the info score is completely dependent upon the (inverse) effective distance (for $\kappa = 0$) or else purely on the extant edge weights ($\kappa = 1$). κ is thus a proxy for how conservative agents are in adjusting their nodes, with higher values reflecting a greater preference to maintain extant connexions. The higher the edge conservation coefficient, the more inertia an edge has proportional to its standing weight, which is intended as a proxy for the variety of psychological heuristics or environmental factors preventing rapid adaptation. In effect, larger values of κ lead to less adaptability, which in the cases of random source seeding helps preserve more globally efficient network structure, while for consistent source seeding prevents the development of specialized local structures, especially in the case of varied initial edge values.

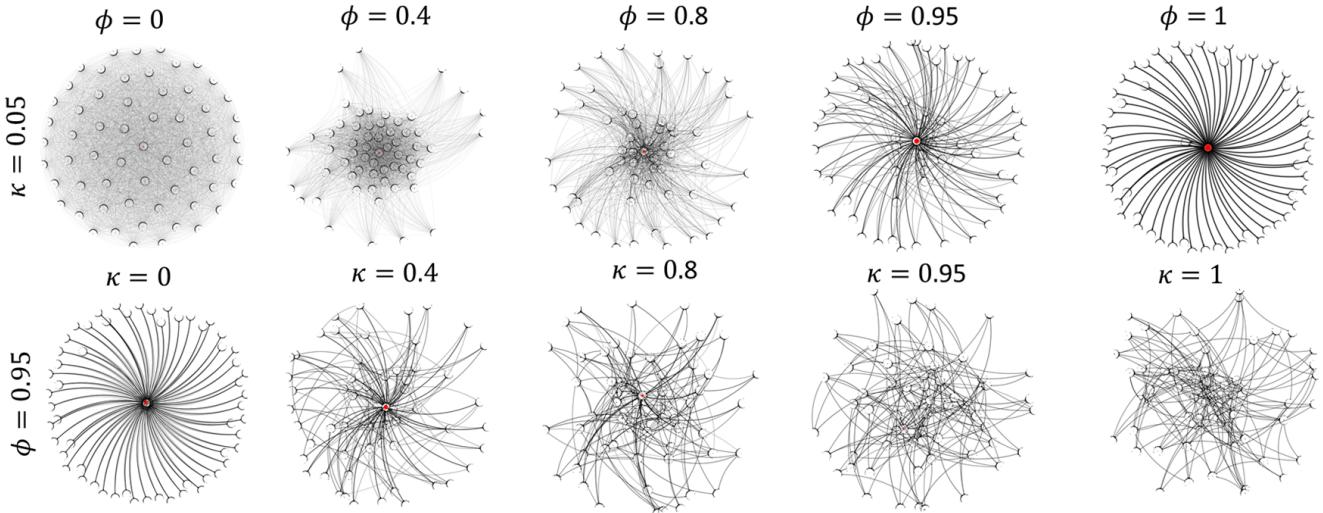


Figure 5.5: Final network structures for a complete graph with constant seeding, evolving through $\phi - \kappa$ space, illustrating the emergent formation of hierarchy as $\phi \rightarrow 1$ and the resistance thereto as $\kappa \rightarrow 1$. This base case of a complete network with constant seeding, where agents are given a maximal set of connections and a consistent objective, best shows the structural results of agent's local optimizations, where nearly every other configuration exhibits intermediate behavior constrained by the model's configuration.

Selectivity ϕ determines the fraction of edges which a node reinforces after every run, and thus is a potential proxy for either the rapidity or certainty of the agent's response. Values of selectivity $\phi > 1/\langle k \rangle$ lead to more marked effects as then selectivity is more discerning than the agent's default ability, enforcing a reduced selection of edges to reinforce. Selectivity has the clearest effect on network structure, as the selectivity parameter determines the number of edges which are consistently reinforced; which for

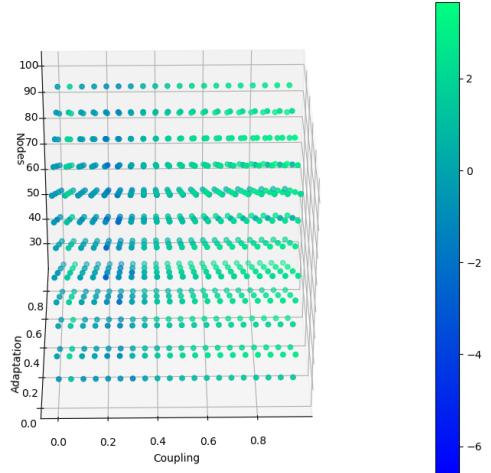


Figure 5.4: Averaged global effective distance differences over n, κ and ϕ

selectivity $1 - \phi > 1/k$ forces a single edge to be reinforced every run, which may rapidly lead to a single edge having all the weight, which when paired with edge weight factoring into the reinforcement ($\kappa > 0$) may lead to a node 'stuck' reinforcing a single edge. Naturally, high values of selectivity thus promote potentially problematic regimes of perpetual reinforcement and inadaptability, though most networks require a value of $\phi > 0$ to progress into more specialized structures.

5.1.5 Null Model

In order to observe the effects of the model constraints *independent* of the effects of reinforcement via simulated diffusion, a *null configuration* was created which preserves the fundament of the model and its dependence on ϕ and κ while eliminating its dependence on effective distance (and thus seeding mechanics). Given the result of primary interest is the emergent network structure via local optimizations to information diffusion, it is necessary to distinguish network structure which arises purely as a result of the mechanics induced via varying selectivity and edge conservation, and that which arises as a function of nodes' optimization processes. The null configuration thus simply replaces the info-score calculation, which previously relied on the random walker effective distance as in equation 3.1, with

$$\text{Info Score}_{ik}(\kappa) = \left(\frac{\mathfrak{D}_{ik}}{\sum_j \mathfrak{D}_{ij}} \right)^{\kappa-1} \cdot \mathbf{A}_{ik}^\kappa \equiv \mathcal{S}_{ik} \longrightarrow \text{Info Score}(\kappa)_{\text{null}} = \mathbf{A}_i \cdot \kappa + \mathbf{R}_i \cdot (1 - \kappa) \quad (5.4)$$

where as initially, κ is *edge conservation*, $\mathfrak{D}_{i;j}$ the i^{th} node's effective distance (from the source node n_j for the directed connection), and $\mathbf{A}_{i;k}$ are the edge values connecting to the i^{th} node from every one of its connections, which are evaluated separately, and \mathbf{R} is a matrix of uniformly random values $\in (0, 1) \subset \mathbb{R}^N$. By this simple change in info-score evaluation, selectivity and edge conservation are allowed to play their intended roles, wherein edge conservation determines the extent to which nodes deviate from their initial values, while selectivity continues to determine which percentile of nodes are rewarded.

Ensemble averaged null runs were conducted, and the same observables analyzed where possible (source-dependent properties were impossible, as null simulations are run without reference to a source), and the results demonstrate that selectivity competes with edge conservation to control individual node edge distribution; the higher the selectivity and the lower the edge conservation, the fewer edges each node reinforces. This relation determines the resultant baseline all to all effective distance measures, with generally worse performance as $\phi \rightarrow 1$, $\kappa \rightarrow 0$.

Examination of the resultant hierarchy coordinates however reveals that while spontaneous hierarchy is noticeable for high selectivity regimes, there is markedly less network hierarchy overall, as is expected due to the lack of a source to centre optimization and implicitly, hierarchy. It is notable that edge conservation has a clearer effect in the absence of the confounding influence of the source, as κ then determines the rate of adaptation, which in the context of uniformly random marginal edge rating probabilities, reduces the influence of selectivity.¹⁷ Thus the null simulations show that though the trends in system parameters of the network generally, such as the convergence of the individual node degree distributions (selectivity forces larger portions of every node's individual edge weight upon small fractions of their edges) and some natural development of hierarchy under constrained randomness¹⁸, the distinct structures which emerge through the model's coupling of information diffusion and (conserved) edge reinforcement are not due to the system constraints.

This independence of the emergent network structure from the model's constraints are readily demonstrated by subtracting the ensemble averaged null-model networks from those obtained through each of the ensemble-averaged configurations, and subsequently re-normalizing and performing the same observable calculations. Should the thusly 'null normalized' metrics remain unchanged, then this provides evidence

¹⁷It is even more notable that this effect holds even in the comparably random case of random seeding, which deprives the network of the ability to relax into an optimal regime through constant source movement.

¹⁸The final figure of [30] shows the accessibility of the entire hierarchy morphospace through natural random variations.

that the emergent network structure is not the result of the model's mechanical constraints, but rather the optimization to diffusive communication.

The null normalization depicted in figure 5.6 shows the effect of diffusive communication even in the

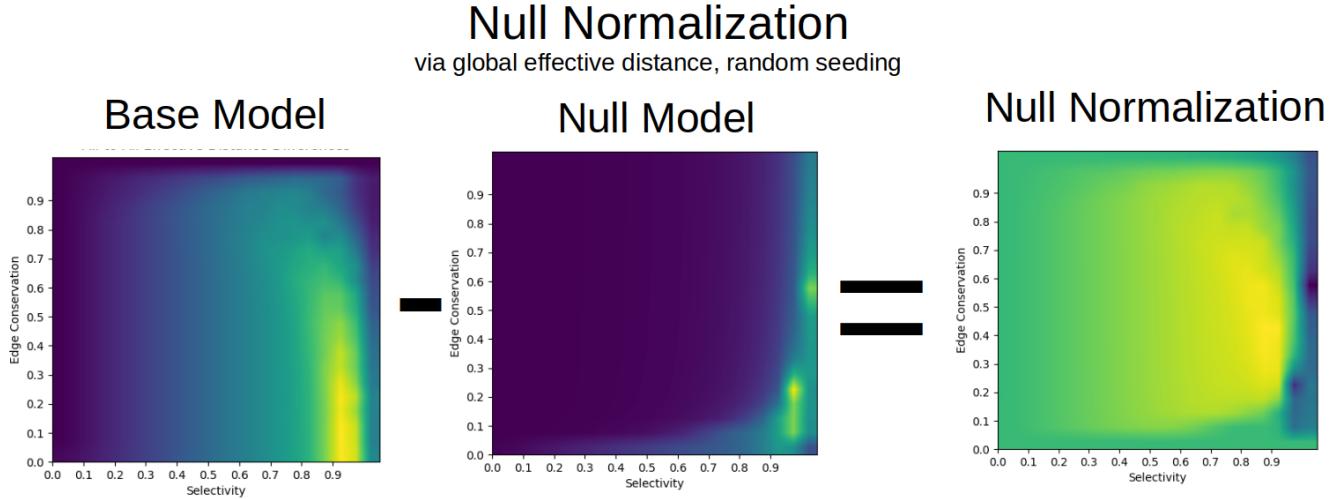


Figure 5.6: Null normalization reveals the effect of local optimization *to diffusive* communication (independent of the model's ϕ, κ constraints) and is given by the difference of ensemble averaged base and null model observable values. Pictured results come from ensemble simulations of 25 simulations, 60 nodes, $\delta = 10$, 250 runs per simulation, $\rho_{\langle k \rangle} = 1$ (fully connected graph) intervals of $\kappa, \phi = 0.05$, bi-linear interpolation.

context of random seeding, wherein agents are unable to adapt to the fluctuating source, but still form a structure which leads to a less effective global optima than the random, null model. This shows a marked tendency towards more solidified structure due to local optimization to diffusive communication, despite the reduction in *global* communication efficiency that arises therefrom. This structural specialization however leads to a considerable reduction in the *average* shortest path between all nodes for $\rho_{\langle k \rangle} = 1$, as one would expect from the development of more solidified structure.

While it is true that there are (combinatorically) many distinct network structures which produce some of the same observable results (e.g. for all to all effective distance), and thus subtracting one individual network from another may not yield the clearest evidence of their respective independence, as the networks are initialized the same and averages are taken over an ensemble, the results are significant. If the networks compared were not identically initialized however, one would need an ensemble proportional to the number of possible network configurations which could produce similar metric results for the difference to be significant. This forces individual network runs and their respective observables to be averaged over the course of the ensemble, and not take the difference of the ensemble averaged null and default configurations, or alternatively simply take the difference of their normalized observable outputs.

5.2 Model Behavior

The central result observed from local optimization to diffusive communication, confirmed through all enumerated observables and robust against all modifications, is the formation of a star-like structure wherein all peripheral nodes connect to emergent hubs centered around the most consistent source. This result requires sufficiently consistent (nearly constant) seeding, high selectivity and low edge conservation, while virtually every other modification leads to natural deviation from this core behavior. Critically, this behavior is *not* observed in the null case, and thus is uniquely the result of optimization according to the

diffusive communication modeled, and not the constraints of the model. That this behavior is observed only for sufficiently high selectivity and low edge conservation suggests that this central, emergent behavior is especially sensitive to its environmental constraints, if not initial network. The clearest evidence of this behavior is seen through observation of the network evolution directly¹⁹, however examination of the associated weighted degree distribution's bifurcation²⁰ reveals the development of this 'star-like' structure, as in figure 5.13.

Local optimizations to a consistent information source in a diffusive communication regime leads to

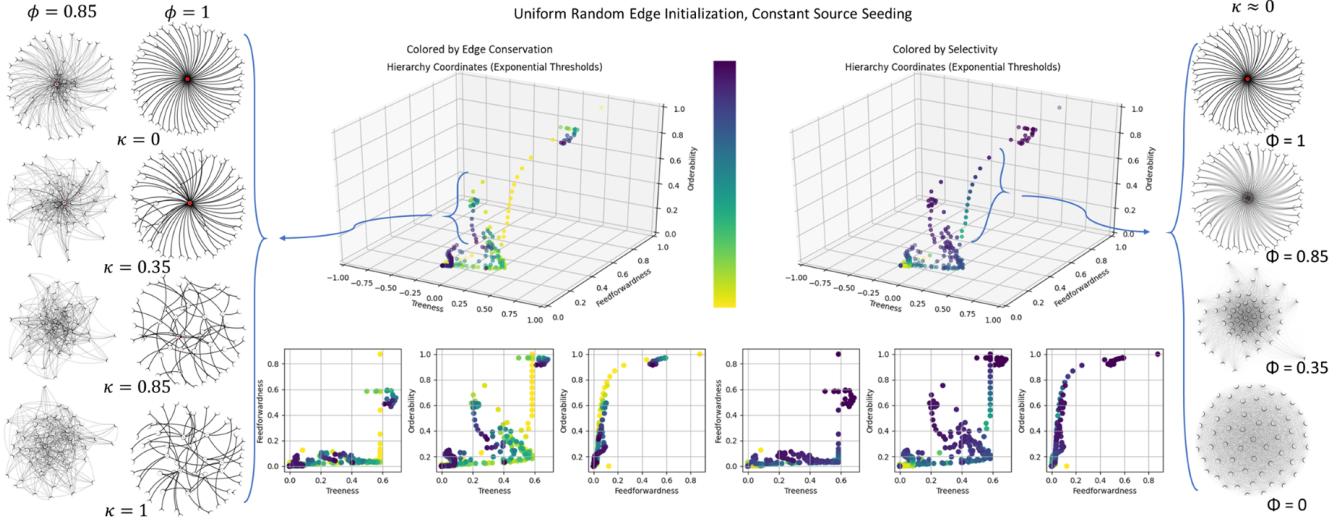


Figure 5.7: Hierarchy morphospace for constant seeding on complete networks, pictured with associated final network structure through the parameter space of edge conservation (left) and selectivity (right). Higher edge conservation values leaves the network unable to adapt to the constant source and thus with comparably low values of all hierarchy coordinates, particularly orderability, while higher selectivity allows for the network to adopt some hierarchy in its reinforcement of a narrower set of edges, thus the lower hierarchy coordinates are the result of very high κ or low ϕ . Hierarchy coordinates are plotted from 25 simulation ensemble averages, graphs are from single simulations.

a centralized structure, with peripheral agents shedding their auxiliary connections in favor of greater proximity to the source. Though this is most apparent for connected networks with sufficiently selective agents, this behavior is observed to greater or lesser extents as the model mechanics permit, as may be seen through investigation of the model's associated configurations, null results, and perhaps most prominently through the resulting hierarchy coordinates. While this result alone is significant in its implication that crude communication networks are capable of self-organization into hierarchies to maximize the reach of the first step(s) of diffusive communications, there are further implications regarding the effects of the psychological heuristics encapsulated in selectivity ϕ and edge conservation κ .

While selectivity may be interpreted as a simple psychological notion of literally how selective one is in their relations, it may be more readily interpreted in terms of the more common and readily applicable notions of frequency of contact/reinforcement or certainty. If an agent is more certain of the efficacy of a given connection, they are more likely to be more precise in their reinforcement thereof, or alternatively if they are simply not in a position to encounter a given contact with sufficient regularity to reinforce them, they shall necessarily be more selective in their choice thereof. Thus given the emergent correlation

¹⁹Animations of the network evolution, as well as a basic sketch of this convergent behavior may be found with the implementation at the associated github repository. [54]

²⁰The unweighted degree distribution necessarily remains nearly constant, as there is no means of agents adding to their degree, only normalizing out existing connections through continuous support of other edges. The weighted degree distribution however tracks nodes' total incoming edge weights ($\sum_i k_{in} \forall i \in \mathcal{N}$), and thus the development of the star-like structure as nearly all edge weight is shifted to the single constant source node.

between selectivity and hierarchy (see section 5.2.2, and figure 5.7), we may make the rather mundane conclusion that when agents are in a position to better choose; i.e. they are more certain and discerning between their connections²¹, they are more likely to optimize their connections by forming a clear hierarchy centered about a consistent source.²² There is however some subtlety in the trade off between edge conservation and selectivity for inconsistent source seeding; as selectivity increases allowing for the specialization which permits the development of centralized structure, the edge conservation must be high enough to ensure that the same edges are consistently rewarded, while also low enough to ensure that the initial edges are not fixed regardless of the source, as happens at $\kappa = 1$.

Edge conservation may likewise be readily interpreted as its psychological namesake; as favor for existing connections over advantage in proximity to an existing source. This tendency to resist communication optimization likely corresponds to several broader and nearly universally relevant factors, such as faith in the consistency of the source (one is more likely to adjust their communication network should they know the adjustment would continue to be effective), belief in the efficacy of one's existing connections and the implicit reasoning behind their formation *ab initio*, and most importantly, cost of adjustment. In this sense it is natural that edge conservation is negatively correlated with the formation of the signature hierarchy, as it acts as a direct dampening on the adjustments of the network. It is however suspected that in many scenarios, such as inconsistent information sourcing, scale-free network initializations, or coupled seeding regimes²³, edge conservation would come to play a more beneficial role with regards to information diffusion, however all modifications thus far examined through the full spectrum of $\kappa \in (0, 1)$ exhibit only a damped hierarchy formation. Though coupled seeding, the configuration most likely to produce positive correlation between κ and effective distance to source, has yet to be comprehensively tested, the difficulty in exposing the $\kappa - \phi$ space for edge conservation to be advantageous to local communication efficiency is that network adaptations are generally faster than the timescales upon which edge conservation yield better results. There is nonetheless a clear relation between edge conservation and the size and centrality of the final hub formed about a constant source, as follows directly from the dampening of the formation of hierarchy.²⁴

Initial grid searches were conducted through a range of selectivity values $\phi \in (0, 1)$ and edge conservation coefficients $\kappa \in (0, 1)$ with intervals of 0.05 for both, which illustrated trends of exponential dependency on selectivity, so that later grid-searches were performed over the ranges $\phi \in (0.8, 1)$ in intervals of 0.02 and $\kappa \in (0, 0.95)$ in intervals of 0.05.²⁵ Grid searches were themselves performed over a range of model configurations, most notably for varying seeding methods and initial network structure.

Networks were initially initialized with uniform random, scale free and set density $\rho = 5\%$ (i.e. all nodes randomly connected to 3 other nodes in the 60 node network) structures, however as uniform random edge initialization lead to the clearest adaptation, and structural evolution appears more dependant on edge density broadly than the specific structure, later grid-search structures were varied over universal edge density, $\rho_{\langle k \rangle}$. Indeed it is necessary to start with a most general structure; that of a set number of random connections, in order to allow for the network's dynamic adaptation into any structure, rather than initially constraining the network into set structures and expect adaptation within what may be rather stringent constraints on the network's form (as in the case of scale-free edge initialization). Conclusions are thus drawn from examinations of network structures of $k = \{0.1, 0.25, 0.5, 0.75, 1\}$, and compiled

²¹As seen in the associated data compilations [54], this trend is only more relevant, if less apparent in the eventual hierarchy, for incomplete (unconnected) networks for which $k < N$.

²²Even for entirely inconsistent sourcing, connections will be more readily reinforced than without any sourcing at all, as in the null case. (even though no discernible hierarchy is likely to form)

²³Wherein it may be important for the initial hierarchy to be preserved until the source becomes the central node

²⁴The dampening of hierarchy formation consistently leads to better global communication, as specialization for a constant source always comes at the cost of intermediary communication efficiency.

²⁵This reduction of the edge conservation range $\kappa \in (0, 1) \rightarrow (0, 0.95)$ is done in an attempt to eliminate the outlying behavior seen for $\kappa = 1$. The results however displayed outlying behavior for $\kappa = 0.95$, which leads one to suspect a highly exponential edge conservation dependency as well.

graphics consider $k = \{0.25, 0.5, 1\}$.

Grid-Searches are performed for four different seeding mechanisms; constant seeding, random seeding, 10 shifts in seeding, and power law seeding (with $a = 5$). Constant seeding and random seeding may be seen as extremes of the same spectrum, wherein power law and 10 random shift based seeding are natural intermediaries interpolating behavior between the two extremes of seeding consistency.

5.2.1 Grid-Search Heat-maps

Results of the above enumerated grid-searches are summarized in heat-maps for single value observables; All to All (global) effective distance, effective distance to source²⁶, diffusion and routing efficiencies (both remain global measures of communication efficiency [29, 72]), and average shortest path, clustering coefficients, and degree variance.

Broadly, global communication efficiency is optimum for well-distributed edge weights (i.e. all nodes spreading their attention across all contacts they possess) however this is difficult to analytically prove given the nature of the matrix inversion in computing the model's fundamental matrix \mathbf{Z} and the associated random walker effective distance $\mathfrak{D}^{\text{RW}} = -\ln [\mathbf{Z} \cdot \text{Diag}(\mathbf{Z}^{-1})]$. Nonetheless, as all simulation steps adapt the model *away from* this nearly ideal behavior, those parameters forcing more extreme adaptation will lead to less *globally* efficient structures, (see figure 5.10, appendix page 69 for relevant network plots in mapped to parameter space) whereas edge conservation, which dampens the adaptation of the network, will lead to more globally optimal structures, preserving the initial, largely random forms. In the few cases of network optimization which benefits effective distance to source, high edge conservation is likely to hamper the efficacy of the network's adaptation, which when combined with high selectivity, also prevents adaptation into a locally favorable (i.e. re: effective distance to source) network structure.

All to All Effective Distance Differences

There is an exponential relation between the effect of selectivity (through all structures and seeding mechanics) and all to all effective distance, with high selectivity leading to an exponential increase in global effective distance²⁷, which reflects a reduction in global communication efficiency. This is supported via the diffusion and routing efficiencies, which display similar behavior. As discussed earlier, as edge conservation largely plays a role in dampening the effects of selectivity and the naturally emergent behavior of the agent's local optimization, only relatively minor effects arise from edge conservation. There is however a fairly clear progression between edge densities, with $k = N$ having the lowest overall global effective distance, and the *base* values of global effective distance (i.e. results for the lower range of selectivity values) increasing progressively in k . Apart from the clear, smooth transition evident in the case of a complete graph wherein the emergence of a specialized hierarchy centered around a consistent node leads to increasingly poor all to all effective distance (see figure 5.8), there are transitions for graphs with lower edge density. At some point after selectivity exceeds a threshold relative to density, $\phi > 1 - \langle k \rangle$, the network is sufficiently selective to allow for local optimization to produce specialized structure, which naturally leads to less *globally optimal* forms. It is worth noting though that networks specialize after crossing this threshold, which always hurts their global communication efficiency, their specialization only rarely (for consistent seeding, low edge conservation) actually benefits their effective distance to source performance. This is doubtless because for the majority of edge conservation space ($\kappa \gtrapprox 0.2$) the edge

²⁶In calculation of the effective distance to the source, for non-constant seeding mechanisms the average distance of every node to the source is taken, whereas constant seeding allows for a simple difference between the final and initial average distance of all nodes to the source. Though both these averages and differences are relative to each network, there is a further difference between these measures; averages over the course of shifting sources necessarily considers the relaxation periods during which the network adapts to a new source position (as in the case of 10 source shifts) which should skew the averages to be less favorable than a direct comparison between fully adapted networks, as in the case of constant seeding differences.

²⁷Pure examination of the heat maps may be confounded by the non-linear color gradient used (inverse viridis)

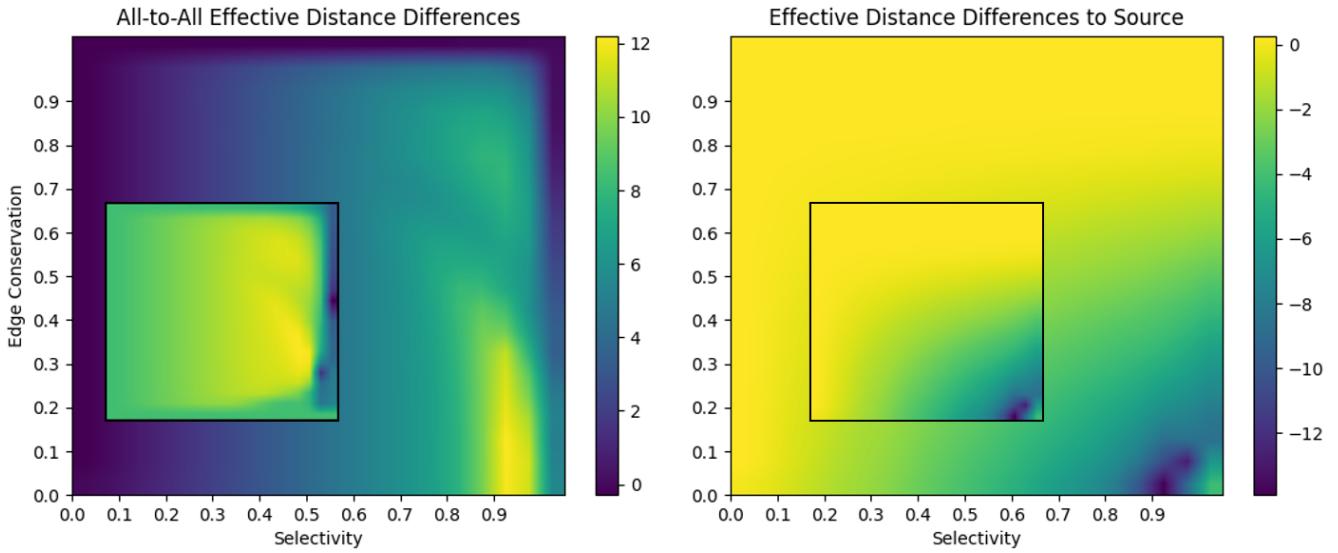


Figure 5.8: Average effective distance differences evaluated for constant seeding, uniform random edge initialization, 25 simulation ensemble, 250 runs per simulation, with bilinear interpolation over selectivity $\phi \in (0, 1)$, edge conservation values $\kappa \in (0, 1)$ and intervals of 0.05. Left: Global (all-to-all) effective distance. Right: Effective distance from source to all other nodes, likewise averaged for both initial and final timesteps, and then the difference is shown. Insets are null normalized versions thereof.

conservation is sufficient to make all attempted specialization futile in effect, though it remains substantial in form for $\phi \gtrapprox 1 - \langle k \rangle$, see figure 5.9

Effective Distance to Source

The marginal benefit of specialization is highly dependent upon edge density: for intermediary $\rho_{\langle k \rangle}$, there is comparably little benefit to centralized communication efficiency, though for complete graphs ($\rho_{\langle k \rangle} = 1$) there is marked benefit in the most extreme cases, as all nodes may be a single diffusion step from the source. In the case of very low edge density, only very high selectivity values will lead to marked shifts in network structure, and adaptations are limited by their initial network structure such that even with perfect global knowledge and optimal rearrangements, the benefit may be negligible as there are no shorter paths to the source. Comparison of these selectivity-dependant transitions as shown via global (top) and source (bottom) effective distances in figure 5.9. Note that the presence of clearly delineated behavior for $\kappa = 1$ for all graphs (and especially noticeable for $\rho_{\langle k \rangle} < 1$) shows the performance of the original configuration, reduced to $\langle k \rangle = (1 - \phi) \cdot \rho_{\langle k \rangle}$ edges, in relation to the 'default' specialization allowed for $\kappa < 1$. Furthermore, as these transitions are entirely absent from the null model with identical constrains, it is clear that these effects are a result of the diffusion process and corresponding local optimization, and not the model's mechanical constraints, though this behavior precisely shows when the constraints of the model prohibit the self-organization that is the central result of this work.

For uniform random edge initialization the point of optimal effective distance to source remains just shy of complete convergence on a star-like graph structure, as is exhibited for $\phi = 1$.²⁸ This suggests that the ideal network structure for a constant source retains some dependence on the strongest, closest connections to the source, as another 'hub of broadcasters'. The mechanics of the simulation force certain structures to form for different values of selectivity and edge conservation, and these structures are clearly inversely effective for various seeding regimes; comparing random seeding to constant seeding reveals that the efficacy of structures seen under constant seeding are inversely effective for random seeding, where the

²⁸This result is nearly assuredly significant, as it is exhibited across ensemble runs of 25 simulations per grid-search point, and furthermore is more readily visible for the reduced range of selectivity $\phi = \{0.8, 1\}$.

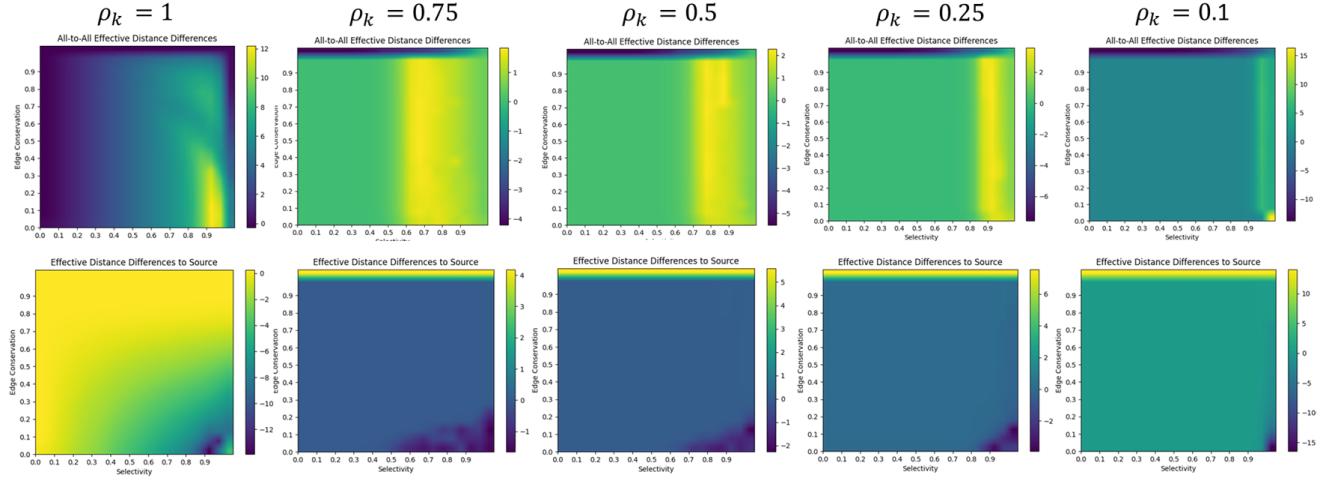


Figure 5.9: Density dependant transitions in selectivity, shown through effective distance differences (all to all effective distance (top) and to source (bottom)) for constant seeding, edge densities $\rho_{\langle k \rangle} \in (1, 0.75, 0.5, 0.25, 0.1)$. The presence of worse global effective distance values for shifting $\phi \propto \rho_{\langle k \rangle}$ indicates the emergence of (globally sub-optimal) network structure, and precisely when model constraints allow for self-organization.

same degree of specialization denies the network ready access to the inconsistent (global) seeding patterns. Power law and '10 shift' seeding provides natural intermediaries between the consistency of constant seeding and random seeding (for which globally optimal solutions perform best), and the results bear this connection out in their apparent interpolation. The vast majority of parameter combinations do not successfully improve effective distance to source, even if specialized, as seen in the post-phase transition areas of figure 5.9. Even a relatively minor set of seeding changes suffices to invert the efficacy of adaptation seen for constant source seeding.²⁹

Diffusion and Routing Efficiencies

Naturally diffusion and routing efficiencies, as global measures³⁰ of communication efficiency are strongly correlated to the global effective distance metric. Indeed, diffusion efficiency is simply a network-normalized version of global effective distance, as detailed in sections 4.3 and 2.3, and routing efficiency is likewise a network-normalized global measure of communication efficiency, though restricted to the shortest paths. Though broadly the behaviors of both diffusion and routing efficiencies are *very* similar to that of all-to-all effective distances, the normalization with respect to other networks allows relative performance, showing that even the worst-performing parameter combinations/configurations lead to network structures which perform 90% capacity relative to the lattice/random networks by which they are normalized. This is significant in its implication that the absolute margins of network performance, upon which global optimization is implicitly based, are relatively narrow, and thus it makes sense that agents with bounded information and explicitly bounded rationality³¹ are unable to achieve absolute gains of < 10% global efficiency, as the signals may be too weak. Specialization into centralized network structures thus arise, as

²⁹Though not pursued in this work, it remains an open question where this line of allowed number of seeding transition is, wherein the network is still able to *successfully* adapt to decrease effective distance to source. Naturally seed adjustments should be rare, and when they do occur, only transfer to the immediate members of the source's hub, so that the network is still allowed to optimize for a given central *hub* of nodes.

³⁰Global measures imply not merely involving all nodes but the relationships between every node to every other; this requires communication through intermediary nodes (i.e. via higher order paths)

³¹Agents have explicitly bounded rationality and information as per the system parameters, $\phi \& \kappa$ (see section 5.1.4)

mandated by the model mechanics, from the absolute gains of individual agents in their *local adaptations* in optimizing performance relative to the source.

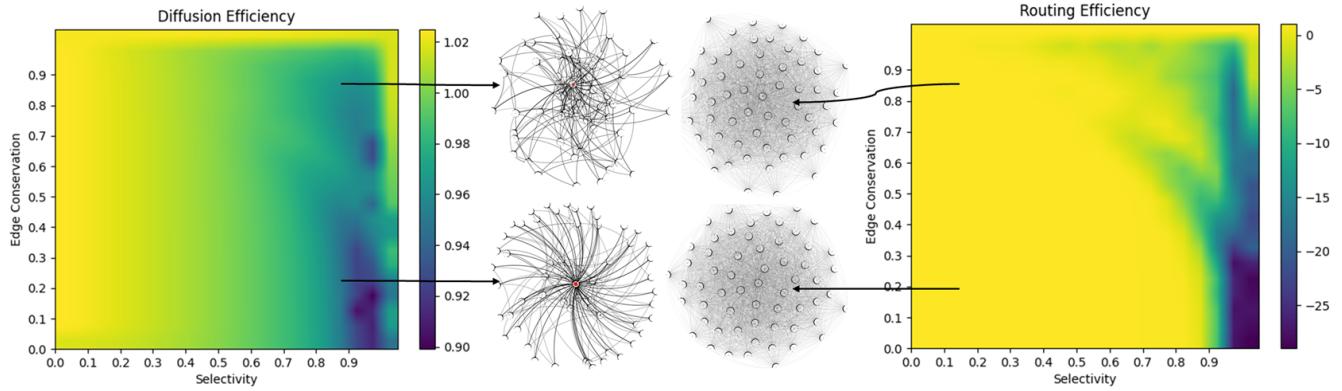


Figure 5.10: Diffusion (left) and routing (right) efficiency heatmaps over the ϕ, κ parameter space for constant seeding, uniform random edge initialization, plotted as heatmaps, with associated networks for each corner shown in the center. Note that both efficiency coordinates are normalized relative to lattice and random networks of equal order, though the normalization does not place the values between 0 and 1 for weighted networks, see sections 4.3 and 7.5.1

Average Cluster Coefficients and Shortest Paths

The averaged cluster coefficients exhibits dependably continuous behavior which helps understand the consistency of underlying network structures, and the reliable effects that parameters ϕ and κ have on network adaptation, particularly evident for the null case, as shown in figure 5.11. The averaged shortest paths however, while likewise exhibiting similarly dependable continuous behavior for the null model, show the effects of the network adaptations when considering non-null model results. Low selectivity preserves the original number of connections, and furthermore inhibits the agent's ability to specialize, which lead to more connected networks, and the resiliently larger clustering coefficients. Notably, random seeding leads to $\approx 50\%$ *more clustering* than seen in the null model, while constant seeding leads to $\approx 50\%$ *less* clustering than seen in the null model, though these results may be dependant upon the network size N .³²

The average shortest path increases as edge density is reduced, as expected, which gives further evidence to the notion that diversification of the same total edge weight across more edges helps improve all to all effective (and direct, shortest path) distances. This result is perhaps best seen in figure 5.12, where one may note the simultaneous increase in the absolute range (lower average shortest paths are seen for smaller ranges of high ϕ) and value of average shortest paths. Though many of the more relevant conclusions regarding the structure and efficiency of the resultant networks are summarized by the purpose built hierarchy and efficiency coordinates, examination of their component observables nonetheless helps affirm the overall behavior and understand the specific trends of the model through ϕ - κ space.

³²Unfortunately average cluster coefficients were not among the parameters initially examined with respect to the network size, though this would likely play a role in the afore mentioned relative effects.

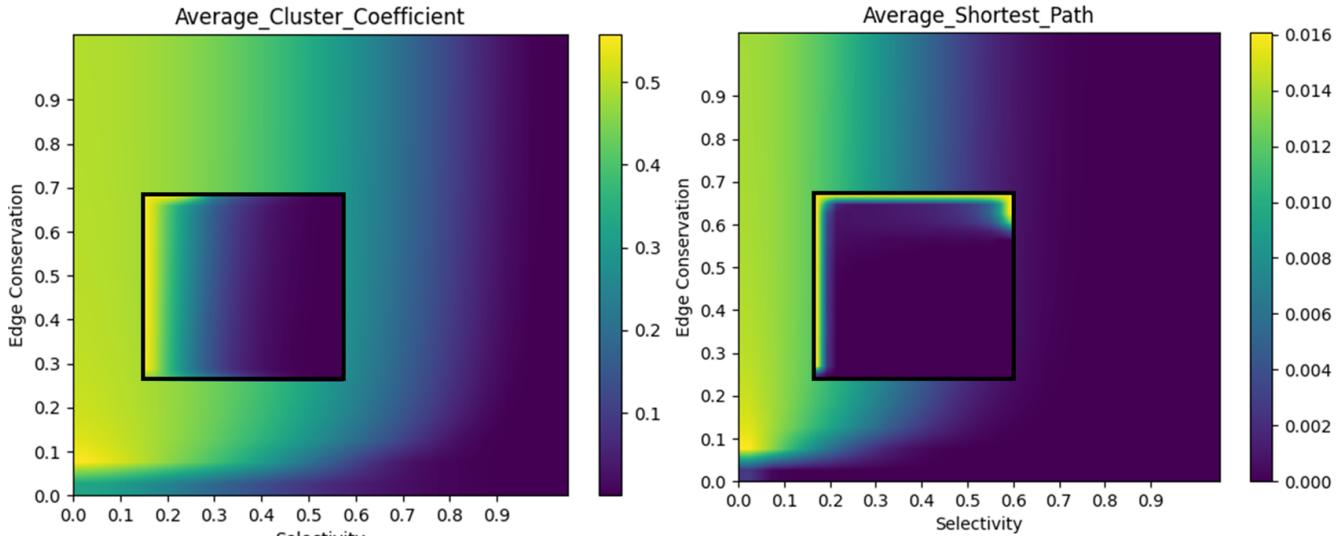


Figure 5.11: Average cluster coefficients and shortest paths of null networks on complete graphs, with insets of the constant seeding versions thereof. There is a consistent, continuous transition for the null case, as expected due to the continuous transition in selectivity and no emergent behavior or corresponding discontinuity.

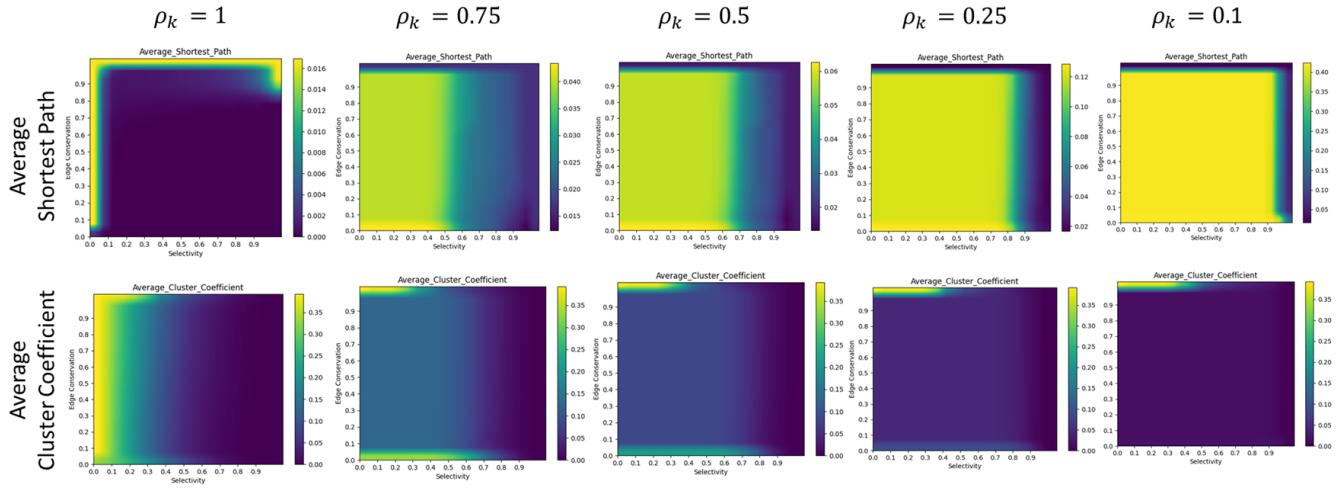


Figure 5.12: Density dependant average cluster coefficient and shortest paths for constant seeding: Note that average shortest paths increase in $\rho_{(k)}$ as expected, despite the conservation of overall edge weight, reinforcing the notion that all to all effective distances corresponds to the diversification of edge weight.

Average Degree Variance

Differences between the values for maximal edge conservation and selectivity and all others are large enough to obscure any more mild variation through the rest of the parameter space, however differences in seeding methodology leads to significant shifts in the high-selectivity degree variance. Consideration of degree distribution histograms for high selectivity and edge conservation regimes shows the bifurcation of the edge distribution from a roughly normally distributed set of edge weights centered around 1 with $\sigma \approx 0.1$ to an increasingly exponential distribution of edge weights with a marked outlying node (or small collection of nodes) with the remaining, increasingly predominant, node weight. This may be seen in figure

5.13, where the number of nodes which possess a collective out degree³³ less than 1 increases exponentially, while there are a few nodes which gain the remaining edge weight. This bifurcation in edge weight

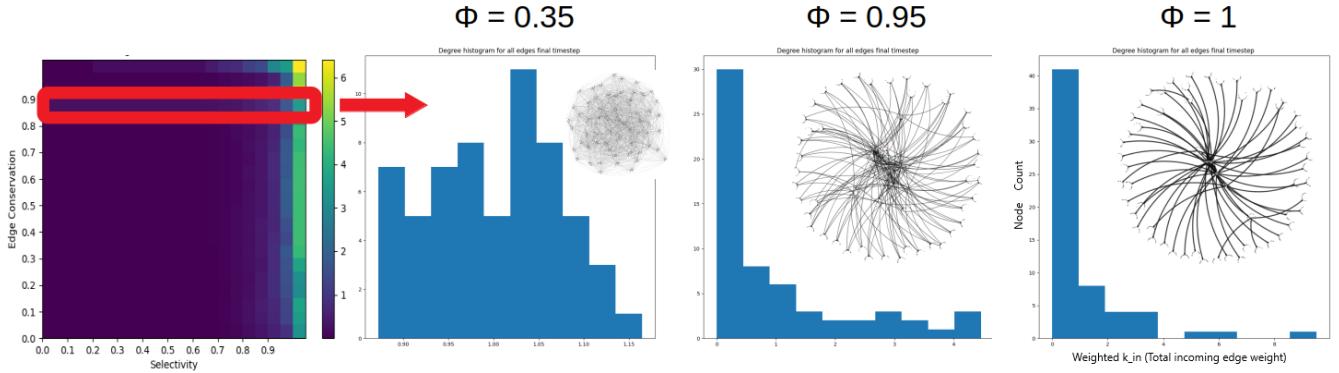


Figure 5.13: Degree progression for power law seeding, edge conservation 0.85, selectivity $\phi \in (0.35, 0.95, 1)$, with sample networks inset. The heat map to the far left depicts the degree variance through the entire parameter range, averaged over 25 runs (as are the degree distributions).

distribution is perhaps the clearest evidence of the emergent hierarchy that forms to optimize diffusive communication from consistent sources, and remains most readily visible in the associated weighted degree distribution histograms, though these are less practical to examine than the degree variance heatmaps. Examination of the weighted degree distribution shows that even for non-constant, yet consistent seeding (as for power law seeding, pictured in figure 5.13) the same formation of centralized network structure occurs, even if slightly more difficult to determine through other metrics. The constant seeding case leads to both the most distinct structures and greatest differences in average degree variance, by nearly an order of magnitude, though this may be node number specific.

As the presented configurations conserve incoming degree weight, nearly all resultant networks are hierarchical³⁴, with treeness > 0 . Orderability increases in selectivity, though the exact correlation depends entirely on degree density; for lower degree density, selectivity must be proportionally higher to take effect, as discussed in section 5.1.4. This effect is to be expected, as higher selectivity leads to a reduced set of edges appearing in the final network structure, which is also more likely to be star-like (and thus more orderable) should the simulation attempt to adapt to a (quasi-) consistent source. Just as the majority of the parameter space displayed similar behavior in the observable in question, save for a relatively limited portion of the higher selectivity-valued parameter space, so there are two distinct regions of the hierarchy morphospace occupied by the results of the model: a value of relatively high orderability, feedforwardness and treeness, and another, larger segment of results spanning the orderability space, and resting at relatively low feedforwardness and intermediate treeness values. The networks plotted appear to 'stretch' in a curve towards the $(1, 1, 1)$ coordinate, where rests the perfect star-like structure seen by comparison to the sample network graphs, as highlighted in figure 5.7

5.2.2 Hierarchy Morphospace

Overall, the diversity of network structures produced through the configuration and parameter grid-searches span a considerable portion of the hierarchy morphospace, and there is a marked difference in network structure due to seeding methodology. There is also a marked difference in the value of hierarchy coordinates as derived from exponential vs linear thresholds; this difference coincides with the shift in

³³Out degree is not necessarily conserved, and thus individual nodes may possess collectively any fraction of the total available edge weight.

³⁴The only exceptions to positive treeness occurs for low edge density, $\rho_{(k)} \approx n/4$

degree distribution as shown in figure 5.13 and is generally present for all configurations along the axis of selectivity.

Though not immediately apparent via consideration of the hierarchy coordinates as plotted via exponential thresholds, the linear thresholds reveal the negative correlation between edge conservation and hierarchy coordinates. This is an expected result, as higher edge conservation yields more consistently random (as per the randomly initialized edges) networks, which are, *ipso facto*, without hierarchy. The effect of seeding is overshadowed by the effects of the parameters themselves, however as variation in seeding prevents or assists (via more or less randomized seeding, respectively) the development of the star-like structure of perfect hierarchy, so the resultant hierarchy coordinates appear more or less well organized in their approach towards a perfectly centralized mono-layer hierarchy³⁵. As with the observable heatmaps, it is readily apparent that constant and random seeding are opposite ends of the spectrum of source consistency, and thus network adaptability. This is born out in the apparent interpolation in network structure that 10 shifts and power law seeding provide, yielding hierarchy coordinates which are between the groupings seen by constant and random seeding.

Comparison of hierarchy coordinates derived from linear vs exponential thresholds helps to distinguish the resultant network structures: If a linear thresholds produces a wider range of hierarchy coordinates, the associated networks possess a more even edge weight distribution, if exponential thresholds yield points mapping to a wider range of the morphospace³⁶, the associated network is likely to possess a bifurcated edge weight distribution. (see figure 5.7) What may appear an amorphous blob of data in one mapping may appear a trend in another, as occurs for the expansion of the low-feedforwardness values of the exponential threshold coordinates in linear thresholds, as in figure 5.14. The entire 3D hierarchy morphospace for all $\kappa - \phi$ combinations is plotted on the top of figure 5.14 to provide for more ready comparison to known real-world network's, as in figure 5.15, while three 2D profile slices thereof are provided, with adjusted axis ranges for ready comparison. Details regarding the adaption of hierarchy coordinates to weighted networks may be found in the appendix, section 7.5.1. As with all tested configurations, compiled versions of all results may be found in the associated git repository, [54].

Comparison of the model's resultant hierarchy coordinates consistently reveal similar regimes to those seen in real world networks, as may be readily confirmed via consultation of 5.15, however this is partially necessarily so, as portions of the hierarchy morphospace are theoretically inaccessible. [30] See figures 4.6 and 4.3 regarding which aspects of the hierarchy morphospace are theoretically impossible to achieve, and [30] regarding the accessibility of the entire morphospace via random variations. Overall the evolution through selectivity of all network configurations leads to a lower average degree, and a more pronounced hierarchy, as is consistent with both the degree (color scale) and general trend of the empirical networks graphed by Corominas et al. in figure 5.15. The resulting hierarchy coordinates nonetheless have distinctly higher treeness than average and than many empirical networks, suggesting that there is convergence to the underlying, highly hierarchical network as the cutoff threshold (see section 7.5.1) increases, and that this convergence to a nearly perfectly hierarchical network (save source's own edges) is near universal. Hierarchy coordinates so constructed of threshold on edge weights thus reveal the behavior underlying adjustments which adjust weighted networks; this may be of wider application, for all networks which are relatively static in their edge distribution (i.e. resist rewiring) and yet adjust the weight of their edge consistently, such as social networks relatively limited timescales.³⁷

Though there may be groups within the empirical networks plotted in figure 5.15 which are spread about the hierarchy space, it is nearly impossible to ascribe any direct relation between any empirical network and the results of the model. However, given the hierarchy space common to both much of the model's

³⁵A mono-layer hierarchy is one with *only* the maximal and minimal set of nodes, as per that pictured in the top right ($\phi = 1$) of figure 5.5. In this case the network possess a trivial maximal hierarchy. (see [51], section 4.2)

³⁶As for the base case of uniform random edge initialization (complete network) and constant seeding.

³⁷This makes the relatively safe assumption that people more readily change the status of their existing connections faster than they make new contacts.

results and empirical networks, one may be able to ascribe the resulting real world networks an evolution consistent with that observed for the relevant model configuration.

5.2.3 Efficiency Morphospace

The results of the efficiency morphospace are individually better examined via their heatmaps, (section 5.2.1) as their positive correlation and respective ranges are best seen in their morphospace plots. Though the respective ranges of diffusion efficiency (centered about one, $\sigma \approx 0.02$) and routing efficiency (also centered about 1, $\sigma \approx 0.05$, though with significantly greater outliers) differ, likely in part due to the imperfect adaptation to weighted networks, they present a clear positive correlation for all but the highest selectivity values. This behavior is expected as both diffusion and routing efficiency are global efficiency metrics; generally, as networks evolve to possess shorter (longer) average paths between points, their shortest possible paths are also reduced (increased). In the edge cases of high selectivity however, the 'star-like' super centralized structures emerge, which prohibit nodes which themselves are not connected to the central source node from possessing a short path, thus increasing overall routing efficiency. Thus, though high-selectivity networks are optimized for information distribution from the central source node, their global communication efficiency, especially when considered exclusively along their shortest possible paths, increases exponentially, as born out in the outlying behavior of the routing efficiency morphospace.

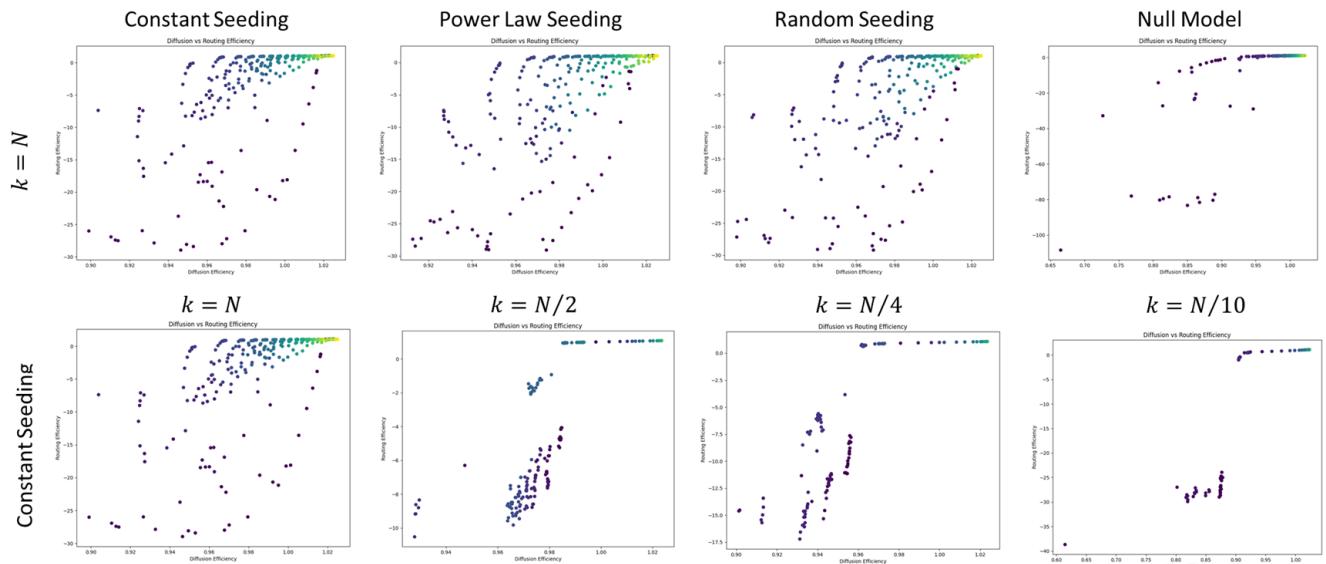


Figure 5.16: Top: Efficiency morphospaces for complete graphs and (from left) constant, power law and random seeding, with the null model results on the far right. Bottom: Efficiency morphospaces for constant seeding, varying edge densities. Note that the differing scales of the axis (especially for $E_{routing}$) make comparison in this overview difficult, thus further results with outliers eliminated are provided in the compilations. [54]

A potentially more interesting pair of parameters to plot together would be the effective distance to source and all to all effective distance, given their imperfect negative correlation, however the trends exhibited in the efficiency morphospace as proposed by [72, 31] show the distinct dependence of final network form on its parameters. When edge conservation and selectivity coloring are compared directly, it becomes apparent that not only is there a continuous shift as suggested by the bilinearly interpolated heatmaps, but each parameter shift produces a distinct trend line in efficiency morphospace: shifts in κ lead to increases in routing efficiency, while ϕ is highly negatively correlated with diffusion efficiency, save for outliers with $\kappa \approx 1$. Thus for randomly initialized networks, edge conservation, while sabotaging the emergent source-centered self organization helps preserve parallelized communication efficiency.

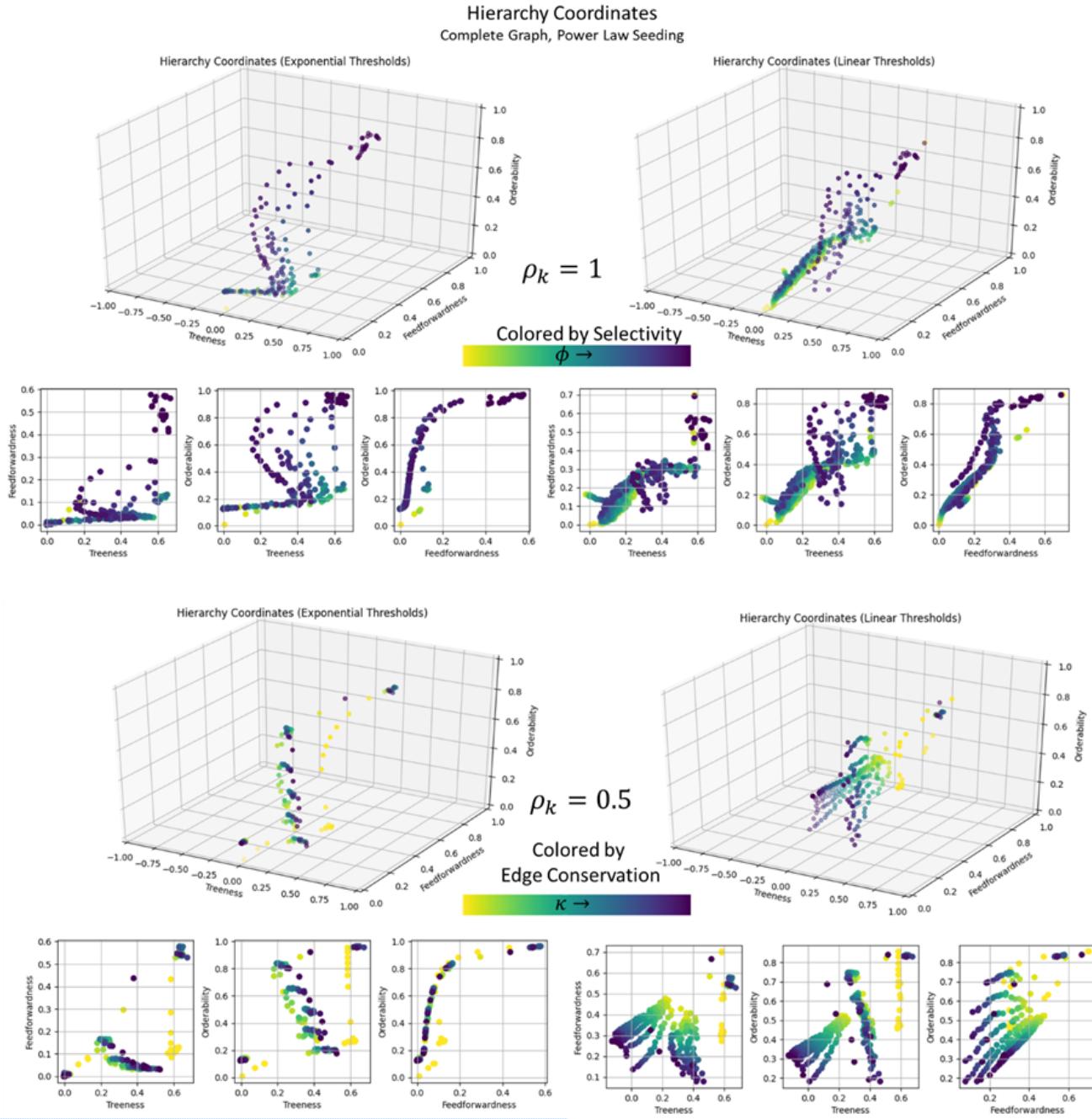


Figure 5.14: Hierarchy coordinates as derived from linear vs exponential thresholds; as hierarchy coordinates derived from exponential thresholds will detect variations in structure evident in more bifurcated weighted degree distributions and hierarchy coordinates derived of linear thresholds may better discern hierarchy between more even weighted degree distributions^a, and thus comparison of the two methods provides further insight into the network structure. The lower selection displays the effect edge density has on hierarchy coordinates, and is graphed using edge conservation as its color code to allow for ready comparison; by examining both color schemes, one may observe the trends that each produce; for instance, the convergence of smaller (lighter) κ values to higher hierarchy values in distinct rows through the bottom left shows how edge conservation resists the formation of structure, while the same plot seen with selectivity coloring, reveals that each rainbow- κ line is a distinct ϕ value, trending towards higher hierarchy with higher selectivity. (most distinctly increasing orderability as $\phi \rightarrow 1$)

^aIn the case of more evenly distributed edge weight, linear thresholds will then not *overweight* the initial network as each threshold will eliminate similar number of edges, and vis-versa.

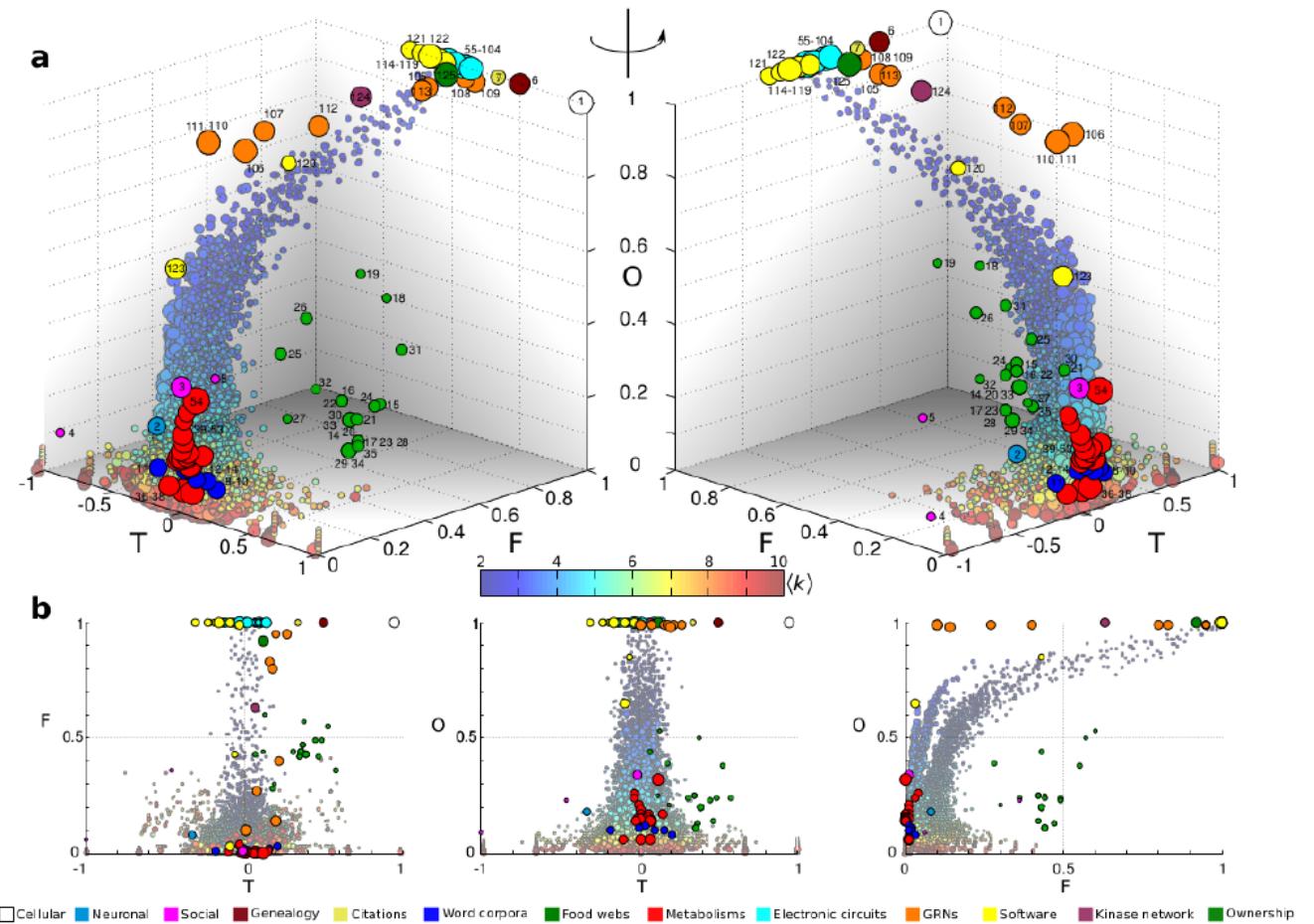


Figure 5.15: Real networks graphed in hierarchy coordinates, color coding according to average degree $\langle k \rangle$. The hierarchy space occupied by the results of the Source to Structure model is the same as many known real world networks, suggesting that model's adaptation process may realistically depict the behavior of the real-world systems which lead to the real world network graphed in this figure. From the supplementary material of [30] as figure 24.

Chapter 6

Conclusion

Seeking to add to our collective understanding of how communication dynamics leads to the formation of diverse networks, this thesis developed a novel model wherein agents permitted only local, conservative adjustments to optimize their diffusive communication efficiency via their existing connections, developed independently into a variety of markedly hierarchical networks. The diverse results which emerged from various model configurations suggest, in combination with their consistent underlying behavior, that the model may effectively represent contributing factors to the known diversity of networks. By varying the model's selectivity and edge conservation parameters, the model is able to explore a range of communications environments which while at their extremes may be entirely theoretical, may also suggest the difference between digital and analogue signal speeds and accessibility. The most pronounced results of the model occur for precisely the environment most readily reflected by digitization, that of low communication costs, and thus low edge conservation. This suggests that those factors which constitute agent selectivity delineate between the development of highly hierarchical networks centered about centralized information sources and their original non-hierarchical forms.

Though highly abstracted and thus otherwise comparably free to interpretation and application, four assumptions constitute the core of the model and thus the applicability of its results: conserved attention, no formation of new edges, a diffusive communication regime, and agent's limited local knowledge. As the conservation of attention, or constant outgoing edge weight, is nearly universally applicable (though the distribution of total attention across all agents may vary) for all real social networks, this does not constitute a great constraint upon interpretation. The static nature of the networks in question is perhaps the greatest limitation in the model's immediate applicability to real world network evolution, as most of the social networks which this model was largely constructed to represent rewire their connections regularly. Interpretation is thus shifted when comparing to rapidly, continuously rewiring networks to the short time-scale version of thereof, or alternatively as long-time averaged connections for incomplete graphs maintaining a constant *average* degree. The presumption of a diffusive communication mechanism is reflected in a wide variety of real world network behavior, as it is induced through the network structure itself: first hitting times for multi-step connections correspond to the number of paths between them, and thus even for a variety of edge weights, a diffusive communication regime often effectively represents the rate and manner of network communication. Limiting agents to their local network in both optimization and knowledge is likewise no great constraint upon the generality of the model in its real world application, as though every agent, conventionally defined, only may modify their own connections, the constraint forcing only local information to do so is likewise ubiquitous. This does not however account for the variety of other information sources and impacts which are ideally subsumed into the effects of the edge conservation parameter.

After introducing the theoretical fundamentals of network science and reviewing the derivation of the random walker effective distance from the absorbing Markov matrix, we developed the Source to Structure model, which served as the focus of this thesis. The description of the Source to Structure model considered its iterative components: Network initialization, seeding, information diffusion, and subsequent agents' mutual evaluation and diffusion, before these same aspects were considered in light of their possible forms and real-world analogues in section 5.1. A brief description of common network science observables was followed by a description of hierarchy and efficiency morphospaces, which allow for the comprehensive measurement of network structure and efficiency, respectively. These observables were subsequently used to detail the model's multivariate behavior across the variety of configurations and consistent parameter space introduced previously.

The development of distinct hierarchy for consistent information sources and sufficiently selective agents demonstrates the organic emergence of hierarchy in response to a consistently desirable agent or concentrated set of agents. The natural negative corollary to this central result is likewise significant: that for insufficiently selective agents, or else agents otherwise confounded, even a consistently popular set of agents will not suffice to establish hierarchy centered around themselves. While this result reveals a smooth transition into more centralized network structure for complete graphs, less connected graphs possess a phase transition in selectivity for $\phi > 1 - \langle k \rangle / N$, whereafter the network may centralize, though only establishing clear hierarchy when edge conservation is low enough. Though overall communication efficiency with respect to the information source may not be clearly improved until a hierarchy is formed, the newly adapted hierarchy coordinates for weighted networks reveal the growing trend towards the same hierarchical structure as is eventually readily seen through direct network plots and observable performance. These simply adapted hierarchy coordinates may prove of use in revealing the trends in hierarchy for especially static, weighted network evolutions. Throughout the course of the investigation it had consistently been clear that there exists a near inverse relationship between the effective specialization of a network for optimization for a single information source, and all other mutual connections, which follows as a direct result from the conservation of attention assumed throughout. This relationship is likely born out through many real world networks, especially as the individual gains of specialization are necessarily the driving factor behind hierarchical specialization. Thus the widely applicable notion of diffusive communication through a network may now be understood to give rise to hierarchical structure when there are local network adaptations towards a constant information source.

Appendix

7.1 List of Variables

Symbol	Description
\mathcal{G}	Graph
\mathcal{E}	Set of edges (lines, links)
\mathcal{N}, n_i	Set of nodes (vertices, points), a node in the set; $n_i \in \mathcal{N} \forall i$
N	The number of nodes of a graph, $ \mathcal{N} $
$\omega(s \rightarrow t)$	walk from <i>source</i> node s to <i>target</i> node t ($t, s \in \mathcal{N}$)
π_{ij}	path between nodes $i \rightarrow j$
\mathbf{A}	Adjacency matrix (binary)
$\mathcal{A}, \mathcal{A}_{i,j}$	Weighted adjacency matrix
k_i	Unweighted node degree
$\mathbf{T}_{i,j}$	Markov Transition Matrix, which <i>for the model</i> equates to the normalized weighted adjacency matrix
\mathbf{v}_i	The i^{th} row of the adjacency matrix \mathcal{A}
$\mathbf{p}_i(t)$	Probability that an ensemble of random walkers is on node i after $t \in \mathbb{N}$ steps
$k, \langle k \rangle$	Edge Degree, average edge degree.

7.2 Data and Code Availability

All compiled results are available within the git repository [54], under *data compilations* and *animations*, as well as the associated python implementation of both the model and non-binary graph adapted hierarchy and efficiency coordinates.

7.3 Computational Requirements

Assuming n nodes (max n^2 edges), the model run cycle consists of the following steps, with associated leading computational order:

$$\text{Model complexity} = \underbrace{\text{Source Selection}}_{\mathcal{O}(1)} + \underbrace{\text{RWED}}_{\mathcal{O}(n^{2.x})} + \underbrace{\text{Info-Scores}}_{\mathcal{O}(n^{1.5}.k)} + \underbrace{\text{Edge Reweighting}}_{\mathcal{O}(n)} = \mathcal{O}(n^{2.x})$$

k is the bit count of the related exponent, which for the floating point $1 - \kappa$, κ used should be 32/64, thus reducing the Info score to a leading order of $\mathcal{O}(n^{1.5})$. Overall complexity is then limited by the double matrix inversion of the random walker effective distance evaluation to $\mathcal{O}(n^{2.x})$, where the exact decimal

.*x* is likely determined by the sparsity of the matrix, and the associated algorithm's reduction. Space complexity is not considered.

7.4 Model Design Notes

7.4.1 Model Binaries

Throughout the model there are a variety of binary choices which effect its core run processes, and may greatly shift what the model is able and intended to reflect, and though cursory explorations of the various permutations of these core characteristics have been performed, only the more revealing results found through variation of other parameters will be considered with these binaries shifted in this work. These 'binary' choices allow for a diversity of possible re-interpretations, from reflecting network evolution in response to agents vying to ostracize the source to limitations in communication efficiency rather than limited attention. Unless otherwise specified, the model functions with the following binary configuration: Effective distances are evaluated *from* the source and reinforcement is inversely proportional to edge rewards (nodes are attempting to wire *closer* to the source), incoming edges are conserved, nodes adapt incoming edges, and the network remains undirected.

- Effective distance directionality: Determines if effective distance is calculated *to* or *from* the source.
 - Effective distance and reinforcement correlation: Determines whether effective distance is *positively* or *negatively* correlated with edge reinforcement.
- Edge conservation directionality: Determines if *incoming* or *outgoing* edges are conserved.
- Edge adaptation directionality: Determines whether nodes are allowed to adjust edges which are direct *to* or *from* them.
- Directedness: Determines whether reciprocal connections are averaged after every run (making the network undirected).

Most significant of the binary choices is the direction in which effective distances are calculated; does information propagate *from* the source or *to* the source, which when combined with a corresponding reversal in info score determination, determines whether nodes are vying for proximity or distance from the source. The determination of network evolution in the context of conscious, communal ostracization is different than that of conventional disease models such as SIR [77] wherein agents' efforts to avoid infected individuals is subsumed into the overall transmission rate β , as every agent uses their local network to best know how to avoid the source. As there is as yet no methodology of passing the source to its nearest neighbors in an attempt to simulate a contagion, this behavior is better understood as social ostracization, or else reaction to a disease in the timescale where communication and contact proceeds much faster than disease transmission.

Edge conservation directionality allows for the fundamental conserved quantity of the simulation, which is by default (i.e. incoming edges are conserved) presumed a proxy for attention, to be unlimited and instead conserve the attention one can give; something which outside the context of technologically enhanced socialization seems eminently feasible. Note that these two conserved quantities are not necessarily exclusive, and furthermore there could exist various distributions of attention or broadcasting ability through the set of nodes, though such an investigation lies outside the scope of this initial work.

By shifting agents from adapting their incoming edges to adapting their outgoing edges, the interpretation shifts from individual agents catering for the favor of those who are closest to the source to a quasi-game theoretic model whereby agents attempt to adjust their connections to better their own incoming connections from others. Though this would appear impossible in the face of most initial networks' complexity, preliminary results show most networks adapt to globally favorable effective distances.... EDIT according

to further results. The final binary, that of *undirectedness* presumes that agents have partial control over both in and outwardly directed edges, as though every agent may only directly adjust their incoming or outgoing edges, any reciprocated connection will be averaged; thus while a follower agent A may devote their entire attention to a single other agent B, if agent B only mildly reciprocates the connection, then the overall connection of both will be adjusted to the average of both connections after every run. This possibility again suggests necessary coordination between agents beyond strict competition for limited attention resources, and so the investigation into this binary is little investigated in the course of this work.

7.4.2 Niave Diffusion

In the case of computational networks, diffusion is necessarily a discrete effect with the information delivered to one node passing through its outgoing edges to all adjacent nodes and thus onward to all nodes. The niave solution operating within this framework of edge weights between 0 and 1 of simply multiplying the 'information nugget' value by each outgoing edge of the seeded node, and thusly, recursively, passing the information through the network until the transfers are below a certain cutoff level is unfortunately ridden with difficulties. Information would naturally pass from the receiver to back to its source, and though this may be readily resolved, (i.e. all nodes which have transferred information are no longer able to receive or further diffuse information) such a resolution counters the possible effects of many otherwise valid paths¹ and furthermore only a subset of the possible paths information may take from a given source to a given node are considered; after all, in a fully connected network (every vertex connects to every other) it is the relative weights and their paths (i.e. how many hops it takes to the source) which determine the ease of information diffusion, and the efficiency of a given connection.

7.5 Observable Design Notes

7.5.1 Adapting Hierarchy Coordinates to Weighted Networks

In order to adapt the hierarchy coordinates developed in [30] for application to a weighted network, the weighted network was broken into a set of unweighted networks based on either exponentially or linearly distributed thresholds. as seen in figure 7.1. For each threshold, all edges whose value exceeds the threshold remains and forms the basis for conversion into a new unweighted, directed graph. The hierarchy coordinates are then computed as ascribed in the supplementary material of [30] and averaged to form the coordinates for a single parameter combination of a given model configuration.

¹See illustration in Info Propigator pdf attached

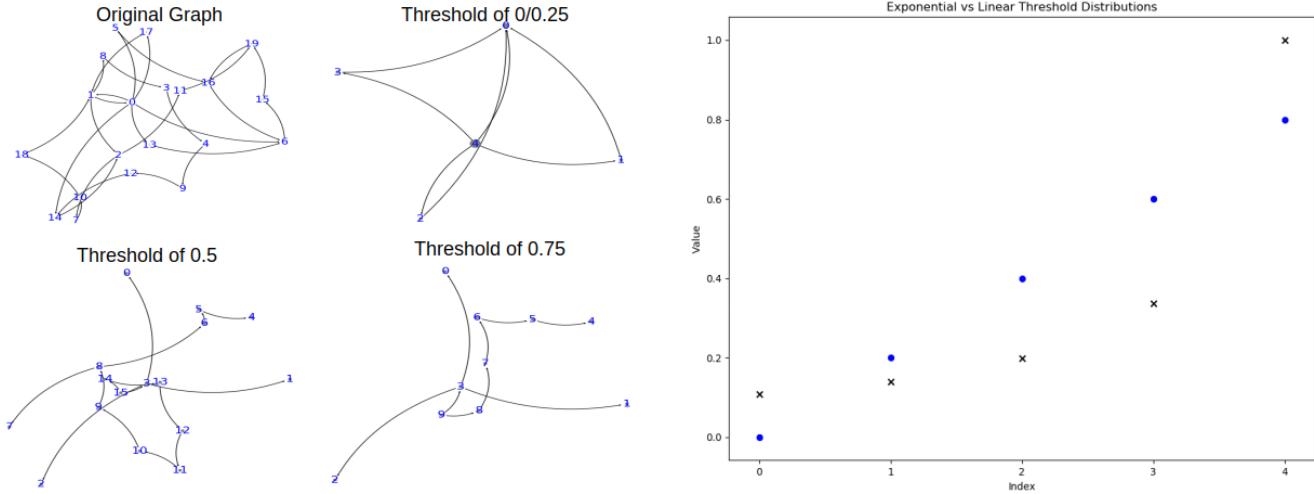


Figure 7.1: Left: Node weighted consolidated directed acyclic graphs, formed from binary network established after threshold's applied. Right: Exponential vs Linear Threshold Distributions

It should be noted that the directed acyclic graphs which result from each threshold-based binary network may have more nodes for higher thresholds, as though there is guaranteed to be as many or fewer (now binary) edges as threshold values increase, this may *increase* the number of nodes which are *not* a component of any cycle. This is perhaps best seen in the transition from the 0/0.25 threshold network's corresponding directed acyclic graph (seen in the top right of the left of figure 7.1) to the 0.5/0.75 threshold versions thereof, wherein the resultant node weighted directed acyclic graphs have lower overall node weight, but more constituent nodes.²

7.5.2 A Note on Observable Ensemble Averages

As many network observables are calculated based purely on network adjacency matrices, it is important to note when an observable evaluated over an ensemble averaged adjacency matrix is the same as the average of observable evaluations over individual adjacency matrices. In the case of node-specific degree distributions, ensemble averaged adjacency matrix will yield the same result as averaging over individual degree-distribution evaluations over every different A in the ensemble. Networks may exhibit structure particular to the model's source selection, and this structure is likely to be 'averaged out' over the course of an ensemble should the corresponding individual simulation's \mathbf{A} be averaged, and the observables evaluated thereupon. This forces the necessity of the computationally more cumbersome evaluation of every observable individually for each simulation, however observable evaluations over network averages are far more likely to resemble the average over their individual evaluations because of the consistency in seeding mechanisms, and the corresponding structure which develops. By means of example, consider the case of constant seeding, which produces the most consistently specialized network structure centered about the constant source node; however, as in every simulation the source node is identical, nearly the same exact structure should result even with variations in edge initialization, as the same node is the constant source. That is, even though we might expect the same network structure to develop from any constant source seeding simulation, because the same exact node (i.e. node number 2, represented via the same row/column index in the adjacency matrix) is chosen every time, the resulting average over adjacency matrices will not be an average over similarly centralized network structures with different central nodes, but the same central node, and thus the ensemble averaged adjacency matrix *does* serve

²Relative number of nodes to total node weight is precisely what orderability, the 'z' hierarchy coordinate, measures. Thus in the example seen in figure 7.1, orderability of the directed acyclic graphs resulting from the 0.5/0.75 thresholds is higher than that of their 0/0.25 threshold based counterparts.

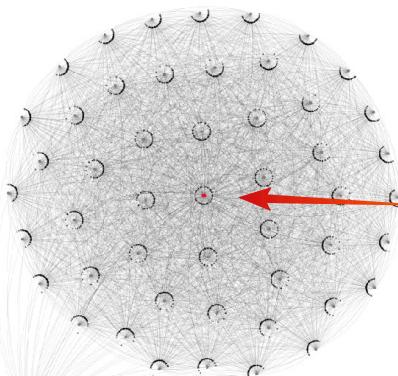
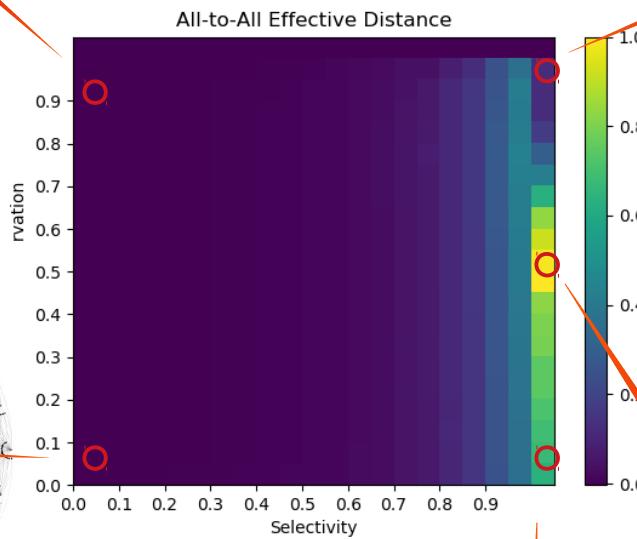
its purpose. Though this is not the case for averages over simulations which witness changing seeding patterns, the inherent randomness in their resultant structures makes the impact of averaging itself less pronounced, though certainly still misrepresenting.

Network Evolution

For Random edge initialization, Constant Seeding, 50 Nodes, Source Reward = 2.6, $\delta = 1$

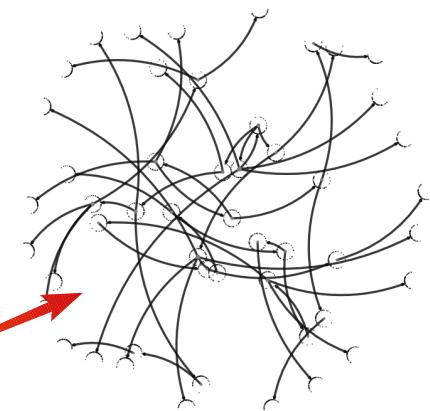
Edge Conservation: 0.9,
Selectivity: 0.05:

As nearly all edges are rewarded according to existent values, the system remains as it was, which by virtue of the high inter connectivity of uniform random edge initialization yields an effectively low global effective distance



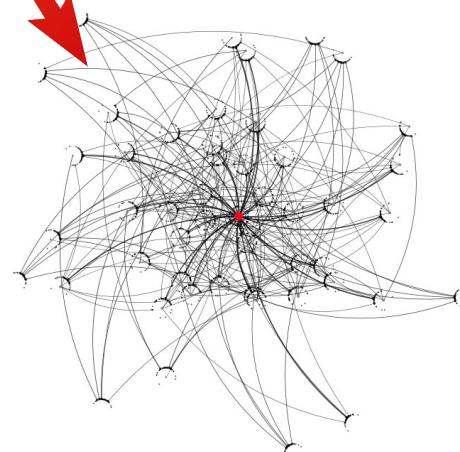
Edge Conservation: 0.05,
Selectivity: 0.05:

Despite all edges being rewarded according to the effective distance from source, the system remains as it was, because all edges are rewarded, and thus the nodes are insufficiently selective to promote a more centralized configuration about the source.



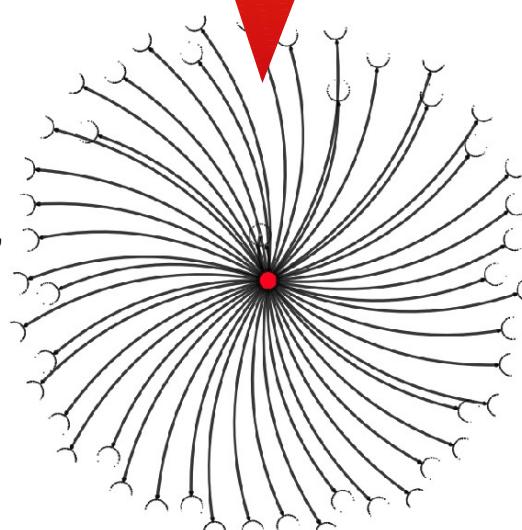
Edge Conservation: 0.95,
Selectivity: 1:

Only one edge per node is rewarded each round, leading to continued reinforcement of whatever the initial edge value was initially strongest, regardless of effective distance to source, due to high edge conservation value.



Edge Conservation: 0.05,
Selectivity: 1:

Only one edge per node is rewarded each round, however as edge conservation is low, the reward is based on effective distance to source, leading to all direct source connections being reinforced, and a resultantly inefficient configuration. (as all connections from the source are reinforced, but not intermediate connections or any to the source).



Edge Conservation:
0.5,
Selectivity: 0.9:

Resting between extremes above and below, the resultant configuration is maximally inefficient.

Bibliography

- [1] Michael Kremer. “Population growth and technological change: One million BC to 1990”. In: *The Quarterly Journal of Economics* 108.3 (1993), pp. 681–716.
- [2] Chris Anderson. “The end of theory: The data deluge makes the scientific method obsolete”. In: *Wired magazine* 16.7 (2008), pp. 16–07.
- [3] Alyssa Friend Wise and David Williamson Shaffer. “Why theory matters more than ever in the age of big data.” In: *Journal of Learning Analytics* 2.2 (2015), pp. 5–13.
- [4] Ahmed Elragal and Ralf Klischewski. “Theory-driven or process-driven prediction? Epistemological challenges of big data analytics”. In: *Journal of Big Data* 4.1 (2017), pp. 1–20.
- [5] Melanie Mitchell. *Complexity: A guided tour*. Oxford University Press, 2009.
- [6] V Vemuri and J William Schmidt. “Modeling of complex systems: an introduction”. In: (1978).
- [7] Brian Castellani. *map of the Complexity Sciences*. https://www.art-sciencefactory.com/complexity-map_feb09.html. 2018.
- [8] William H. Whyte Jr. *Groupthink: Fortune Magazine origin*. <https://web.archive.org/web/20100401033524/http://apps.olin.wustl.edu/faculty/macdonald/GroupThink.pdf>. 1971.
- [9] Freeman Dyson. *A meeting with Enrico Fermi*. 2004.
- [10] Jürgen Mayer, Khaled Khairy, and Jonathon Howard. “Drawing an elephant with four complex parameters”. In: *American Journal of Physics* 78.6 (2010), pp. 648–649.
- [11] Johan Ugander et al. “The anatomy of the facebook social graph”. In: *arXiv preprint arXiv:1111.4503* (2011).
- [12] Bjarke Mønsted et al. “Evidence of complex contagion of information in social media: An experiment using Twitter bots”. In: *PLoS one* 12.9 (2017), e0184148.
- [13] Philipp Lorenz-Spreen et al. “Accelerating dynamics of collective attention”. In: *Nature communications* 10.1 (2019), p. 1759.
- [14] Abdullah Almaatouq et al. “Adaptive social networks promote the wisdom of crowds”. In: *Proceedings of the National Academy of Sciences* 117.21 (2020), pp. 11379–11386.
- [15] Fabian Baumann, Igor Sokolov, Melvyn Tyloo, et al. “The Role of Active Leaders in Opinion Formation on Social Networks”. In: *arXiv preprint arXiv:1910.01897* (2019).
- [16] Soroush Vosoughi, Deb Roy, and Sinan Aral. “The spread of true and false news online”. In: *Science* 359.6380 (2018), pp. 1146–1151.
- [17] Mark Ledwich and Anna Zaitsev. “Algorithmic extremism: Examining YouTube’s rabbit hole of radicalization”. In: *arXiv preprint arXiv:1912.11211* (2019).

- [18] Kristina Lerman and Rumi Ghosh. “Information contagion: An empirical study of the spread of news on digg and twitter social networks”. In: *arXiv preprint arXiv:1003.2664* (2010).
- [19] Fabian Baumann et al. “Modeling echo chambers and polarization dynamics in social networks”. In: *arXiv preprint arXiv:1906.12325* (2019).
- [20] Hal Berghel. “Malice domestic: The Cambridge analytica dystopia”. In: *Computer* 5 (2018), pp. 84–89.
- [21] Елена Леонардовна Болдырева, Екатерина Дуйсембина, and Наталья Гришина. “Cambridge Analytica: Ethics And Online Manipulation With Decision-Making Process”. In: *Cambridge Analytica: Ethics And Online Manipulation With Decision-Making Process*. Future Academy, 2018, pp. 91–102.
- [22] Cathy O’neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.
- [23] Alexander J Stewart et al. “Information gerrymandering and undemocratic decisions”. In: *Nature* 573.7772 (2019), pp. 117–121.
- [24] Carl T Bergstrom and Joseph B Bak-Coleman. *Information gerrymandering in social networks skews collective decision-making*. 2019.
- [25] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [26] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.
- [27] Erzsébet Ravasz et al. “Hierarchical organization of modularity in metabolic networks”. In: *science* 297.5586 (2002), pp. 1551–1555.
- [28] Andrea Avena-Koenigsberger et al. “Network morphospace”. In: *Journal of the Royal Society Interface* 12.103 (2015), p. 20140881.
- [29] Bernat Corominas-Murtra et al. “Measuring the hierarchy of feedforward networks”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 21.1 (2011), p. 016108.
- [30] Bernat Corominas-Murtra et al. “On the origins of hierarchy in complex networks”. In: *Proceedings of the National Academy of Sciences* 110.33 (2013), pp. 13316–13321.
- [31] Joaquín Goñi et al. “Exploring the morphospace of communication efficiency in complex networks”. In: *PLoS One* 8.3 (2013), e58070.
- [32] Flavio Iannelli et al. “Effective distances for epidemics spreading on complex networks”. In: *Physical Review E* 95.1 (2017), p. 012313.
- [33] Douglas Guilbeault, Joshua Becker, and Damon Centola. “Complex contagions: A decade in review”. In: *Complex spreading phenomena in social systems*. Springer, 2018, pp. 3–25.
- [34] Aurélien Gautreau, Alain Barrat, and Marc Barthelemy. “Global disease spread: statistics and estimation of arrival times”. In: *Journal of theoretical biology* 251.3 (2008), pp. 509–522.
- [35] Andreas Kohler. “Three Essays in International Economics”. PhD thesis. University of Zurich, 2013.
- [36] Réka Albert and Albert-László Barabási. “Statistical mechanics of complex networks”. In: *Reviews of modern physics* 74.1 (2002), p. 47.
- [37] Dirk Brockmann and Dirk Helbing. “The hidden geometry of complex, network-driven contagion phenomena”. In: *science* 342.6164 (2013), pp. 1337–1342.
- [38] Karl W Deutsch and Walter Isard. “A note on a generalized concept of effective distance”. In: *Behavioral Science* 6.4 (1961), pp. 308–311.
- [39] Nicola Perra et al. “Activity driven modeling of time varying networks”. In: *Scientific reports* 2 (2012), p. 469.

- [40] Alessandro Vespignani and Stefano Zapperi. “How self-organized criticality works: A unified mean-field picture”. In: *Physical Review E* 57.6 (1998), p. 6345.
- [41] Duncan J Watts and Steven H Strogatz. “Collective dynamics of ‘small-world’ networks”. In: *nature* 393.6684 (1998), pp. 440–442.
- [42] Pedro D Manrique et al. “Generalized gelation theory describes onset of online extremist support”. In: *Physical review letters* 121.4 (2018), p. 048301.
- [43] Greg Paperin, David G Green, and Suzanne Sadedin. “Dual-phase evolution in complex adaptive systems”. In: *Journal of the Royal Society Interface* 8.58 (2011), pp. 609–629.
- [44] Giulio Cimini et al. “The statistical physics of real-world networks”. In: *Nature Reviews Physics* 1.1 (2019), pp. 58–71.
- [45] Douglas J Klein and Milan Randić. “Resistance distance”. In: *Journal of mathematical chemistry* 12.1 (1993), pp. 81–95.
- [46] Emile Chappin. “Review of Phase Transitions (Primers in Complex Systems)”. In: (2012).
- [47] Hartmut HK Lentz et al. “Disease spread through animal movements: a static and temporal network analysis of pig trade in Germany”. In: *PloS one* 11.5 (2016), e0155196.
- [48] Duncan J Watts. “A simple model of global cascades on random networks”. In: *Proceedings of the National Academy of Sciences* 99.9 (2002), pp. 5766–5771.
- [49] John G Kemeny and J Laurie Snell. *Markov chains*. Springer-Verlag, New York, 1976.
- [50] Joseph Klafter and Igor M Sokolov. *First steps in random walks: from tools to applications*. Oxford University Press, 2011.
- [51] Dirk Brockmann and Dirk Helbing. *Supplementary Materials: The hidden geometry of complex, network-driven contagion phenomena*. <https://science.sciencemag.org/content/suppl/2013/12/11/342.6164.1337.DC1>. 2013.
- [52] Andreas Koher. “Effective Distances on Complex Networks”. In: (2019). [Online].
- [53] Edsger W Dijkstra et al. “A note on two problems in connexion with graphs”. In: *Numerische mathematik* 1.1 (1959), pp. 269–271.
- [54] Lorenz-Spreen Bromberg. *Network Via Diffusion*. https://github.com/MaxBromberg/Networks_Via_Diffusion. [Online]. 2019.
- [55] Michael L Fredman and Robert Endre Tarjan. “Fibonacci heaps and their uses in improved network optimization algorithms”. In: *Journal of the ACM (JACM)* 34.3 (1987), pp. 596–615.
- [56] Philipp Gert Josef Lorenz-Spreen. “Dynamics of collective attention”. PhD thesis. Technische Universität Berlin, 2019.
- [57] Javier Borge-Holthoefer et al. “The dynamics of information-driven coordination phenomena: A transfer entropy analysis”. In: *Science advances* 2.4 (2016), e1501158.
- [58] Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. “Modeling users’ activity on twitter networks: Validation of dunbar’s number”. In: *PloS one* 6.8 (2011), e22656.
- [59] Natasa Djurdjevac et al. “Random walks on complex modular networks12”. In: *JNAIAM* 6.1-2 (2011), pp. 29–50.
- [60] Sofía Aparicio, Javier Villazón-Terrazas, and Gonzalo Álvarez. “A model for scale-free networks: application to twitter”. In: *Entropy* 17.8 (2015), pp. 5848–5867.
- [61] Alain Barrat et al. “The architecture of complex weighted networks”. In: *Proceedings of the national academy of sciences* 101.11 (2004), pp. 3747–3752.

- [62] Jari Saramäki et al. “Generalizations of the clustering coefficient to weighted complex networks”. In: *Physical Review E* 75.2 (2007), p. 027105.
- [63] Vito Latora and Massimo Marchiori. “Efficient Behavior of Small-World Networks”. In: *Phys. Rev. Lett.* 87 (19 Oct. 2001), p. 198701. DOI: 10.1103/PhysRevLett.87.198701. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.87.198701>.
- [64] Jianxi Luo and Christopher L Magee. “Detecting evolving patterns of self-organizing networks by flow hierarchy measurement”. In: *Complexity* 16.6 (2011), pp. 53–61.
- [65] Julian J McAuley, Luciano da Fontoura Costa, and Tibério S Caetano. “Rich-club phenomenon across complex network hierarchies”. In: *Applied Physics Letters* 91.8 (2007), p. 084103.
- [66] Linton C Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977), pp. 35–41.
- [67] Mahendra Piraveenan, Mikhail Prokopenko, and Liaquat Hossain. “Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks”. In: *PloS one* 8.1 (2013), e53095.
- [68] Phillip Bonacich. “Power and centrality: A family of measures”. In: *American journal of sociology* 92.5 (1987), pp. 1170–1182.
- [69] Linton C Freeman. “Centrality in social networks conceptual clarification”. In: *Social networks* 1.3 (1978), pp. 215–239.
- [70] Carlos Rodriguez-Caso, Bernat Corominas-Murtra, and Ricard V Solé. “On the basic computational structure of gene regulatory networks”. In: *Molecular BioSystems* 5.12 (2009), pp. 1617–1629.
- [71] M Cosentino Lagomarsino, P Jona, and B Bassetti. “Logic backbone of a transcription network”. In: *Physical review letters* 95.15 (2005), p. 158701.
- [72] Vito Latora and Massimo Marchiori. “Efficient behavior of small-world networks”. In: *Physical review letters* 87.19 (2001), p. 198701.
- [73] Robin IM Dunbar. “Neocortex size as a constraint on group size in primates”. In: *Journal of human evolution* 22.6 (1992), pp. 469–493.
- [74] Hector O Zapata and Wayne M Gauthier. *Threshold models in theory and practice*. Tech. rep. 2003.
- [75] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. “Modeling market mechanism with minority game”. In: *Physica A: Statistical Mechanics and its Applications* 276.1-2 (2000), pp. 284–315.
- [76] Sorin Solomon and Peter Richmond. “Power laws of wealth, market order volumes and market returns”. In: *Physica A: Statistical Mechanics and its Applications* 299.1-2 (2001), pp. 188–197.
- [77] Claude Lefèvre. “Stochastic epidemic models for SIR infectious diseases: a brief survey of the recent general theory”. In: *Stochastic processes in epidemic theory*. Springer, 1990, pp. 1–12.