University of Vermont
UVM ScholarWorks

Graduate College Dissertations and Theses | Dissertations and Theses

2020

Measuring and Modeling Information Flow on Social Networks

Tyson Charles Pond
University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

 Part of the Applied Mathematics Commons, Library and Information Science Commons, and the Statistical, Nonlinear, and Soft Matter Physics Commons

Recommended Citation

Pond, Tyson Charles, "Measuring and Modeling Information Flow on Social Networks" (2020). *Graduate College Dissertations and Theses*. 1239.
https://scholarworks.uvm.edu/graddis/1239

This Thesis is brought to you for free and open access by the Dissertations and Theses at UVM ScholarWorks. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of UVM ScholarWorks. For more information, please contact scholarworks@uvm.edu.

Measuring and Modeling Information Flow on Social Networks

A Thesis Presented

by

Tyson Charles Pond

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Mathematics

May, 2020

Defense Date: March 23rd, 2020
Dissertation Examination Committee:

James Bagrow, Ph.D., Advisor
Laurent Hébert-Dufresne, Ph.D., Chairperson
Chris Danforth, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of Graduate College

Abstract

With the rise of social media, researchers have become increasingly interested in understanding how individuals inform, influence, and interact with others in their social network and how the network mediates the flow of information. Previous research on information flow has primarily used models of contagion to study the adoption of a technology, propagation of purchase recommendations, or virality of online activity. Social (or "complex") contagions spread differently than biological ("simple") contagions. A limitation when researchers validate contagion models is that they neglect much of the massive amounts of data now available through online social networks. Here we model a recently proposed information-theoretic approach to measuring the flow of written information in data. We use an idealized generative model for text data – the quoter model – which naturally incorporates this measure. We investigate how network structure impacts information flow and find that the quoter model exhibits interesting features similar to those of complex contagion. Finally, we offer an analytical treatment of the quoter model: we derive approximate calculations and show dependence on model parameters. This thesis gives rise to new hypotheses about the role of the social network in facilitating information flow, which future research can investigate using real-world data.

Citations

Material from this Thesis has been published in the following form:

Pond, T., Magsarjav, S., South, T., Mitchell, L., and Bagrow, J. P.. Complex contagion features without social reinforcement in a model of social information flow. *Entropy*, 22(3):265, 2020.

Acknowledgements

Thank you to my advisor, James Bagrow, for his consistent guidance. Thank you to my committee for offering their time to review this thesis. Thank you to my Australian collaborators: Saranzaya Magsarjav, Tobin South, and Lewis Mitchell. Thank you to my parents for their lifelong support. Thank you to friends in the graduate program. Thank you to the UVM Math Department for funding me during my master's program.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Online social media platforms have seen explosive growth in the recent years. These platforms connect people from all over the world and allow almost instantaneous communication. A report from January 2020 [2] indicates the monthly active users for prominent social media platforms are: Facebook 2.4 billion, YouTube 2 billion, Instagram 1 billion, and Reddit, Twitter, and LinkedIn at 300 million. Consequently, these platforms generate immense amounts of data every day; for example, Twitter sees nearly 500,000 new tweets every minute. In 2013 it was estimated that 90% of all data had been created within two years. Much of this data is publicly accessible to researchers, and thus these platforms provide the ability to examine human behavior at an unprecedented scale.

Along with the rise of social media, there has been increased interest in understanding how individuals inform, influence, and interact with others in their social network and how the network mediates the flow of information. Previous research on information flow has studied the adoption of a technology [3], propagation of purchase recommendations [4], virality of online activity, [5], and communication during emer-

gencies [6]. The study of information flow is important because it allows researchers to predict the spread of social contagions and may further be motivated by potential marketing applications. Modern concerns of information flow include the role and prevalance of: misinformation, bots, echo chambers, and filter bubbles [7–10]. Finally, understanding information flow and social contagion may provide insight as to how biological contagions spread. For example, recent research suggests models for social contagion are useful in modeling biological contagion [11, 12]. Also, including individual behavioral responses to epidemics in these models helps improve predictions [13, 14].

A standard method of measuring information flow on a social network is to treat information as discrete "packets." For example, one might study how information flows on Twitter by tracing the spread of a particular URL or keyword. We could create a quantitative model to predict the spread of the URL or keyword and then validate the model by comparing with real-world data. Such models have traditionally borrowed from epidemiology, treating a response or adoption of a behavior as an infection, and modeling the spread of the *social contagion*. A contagion is typically classified as either a simple contagion or a complex contagion, with biological contagions traditionally being treated as simple and social contagions being treated as complex [15]. With simple contagions, the likelihood of a individual becoming infected increases linearly with the number of infected neighbors. Hence, simple contagions always have the potential to infect an individual upon a single exposure. On the other hand, with complex contagions, the likelihood of being infected increases nonlinearly with the number of infected neighbors. In complex contagion, multiple exposures **from different sources** are required for the contagion to spread. In so-

cial sciences, this mechanism is called *social reinforcement* [16]. This mechanism is intuitive – an individual is more likely to watch a certain TV show if five friends recommend it to them, as opposed to if one friend recommends it to them five times.

The approach of focusing on a proxy for information and applying contagion models circumvents the challenge of defining an objective measure for information, which would ideally encompass all aspects of how humans influence each other. Additionally, it also avoids the challenge of processing and quantifying natural language. The immediate drawback of this approach is that only a small portion of available data is used to measure and model information flow. Furthermore, people may influence each other without prompting a measurable response, i.e. an individual may visit a link they saw on Twitter but not retweet it. In light of these drawbacks, this thesis studies a more intricate approach to measuring and modeling information flow in online social networks which can incorporate all the text data available. Our approach borrows quantitative measures of uncertainty and influence from the field of information theory, which will be reviewed in Chapter 2.

In this thesis, we examine how network properties can affect information flow when taking an information-theoretic perspective, and compare our results to those of traditional contagion models. In particular, we study the quoter model [1], an idealized model for individuals generating text data within social media, and apply information-theoretic estimators to the model text. Using both random graph models and real-world network data, we compare compare and contrast the results of this information theoretic approach with the results of traditional simple and complex contagion models. Interestingly, we find that the quoter model exhibits several phenomena which are characteristic of complex contagion, despite lacking an explicit

social reinforcement mechanism, the key component of complex contagion.

The rest of this thesis is organized as follows. In Chapter 2 we provide a brief review of important concepts from network science, network dynamics, and information theory. We then discuss similar research on information flow. We conclude the chapter by introducing our approach to measuring information flow – the cross-entropy – along with the quoter model. In Chapter 3 we simulate the quoter model on random graph models and real networks to investigate how network structure impacts information flow, as measured by the cross-entropy. We also simulate traditional contagion models and compare with our approach to modeling information flow. In Chapter 4 we provide an analytical treatment of the quoter model. We begin by offering approximate calculations for the the cross-entropy when the network is a single directed link. We also mention a connection between the quoter model and a more well-known model, the voter model. In Chapter 5 we summarize our contributions from Chapter 3 and Chapter 4 and mention possible directions for future work.

Chapter 2

Background

2.1 Network structure

We use the terms network and graph interchangeably. A network, denoted $G = (V, E)$, consists of a set of nodes V and a set of links $E \subseteq V \times V$. Networks are very powerful modeling tools; in a very abstract view, V is a set of objects and E is a set of pairwise relationships between those objects. This thesis explores social networks, where nodes represent people and a link (u, v) indicates that a social tie exists between individual u and individual v.

2.1.1 Basic network statistics

Real-world networks are generally large (many networks of current study consist of millions of nodes) and thus there is a need to summarize the network. Such summary statistics should contain "important" information about the network, and offer a way to compare two networks.

Two immediate network statistics are the number of nodes $N = |V|$ and number of links $M = |E|$. These characterize the size of the network. Note that in comparing two networks G_1, G_2, both N and M should be taken into consideration. For example, if $N_1 \gg N_2$ and G_1, G_2 are equally "dense" then it is not surprising that $M_1 \gg M_2$. This leads to two definitions of density: (i) the average degree $\langle k \rangle \equiv \frac{1}{N} \sum_{u \in V} k_u = 2M/N$ where k_u is the degree – the number of neighboring nodes – of node u (ii) the edge density $M/\binom{N}{2}$. Definition (i) tells us how many links on average each node has while (ii) is a number in $[0, 1]$ telling us the fraction of all possible links which exist.

2.1.2 Properties of complex networks

A complex network is loosely defined as a network which exhibits at least one of the following nontrivial structural properties.

Broad degree distribution In the Erdős-Rényi (ER) random graph model, each link exists independently with probability $0 \leq p \leq 1$. Consequently, the probability that a particular node has exactly k neighbors follows a binomial distribution, $p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$, which is a polynomial in p. However, real-world networks often have a more heterogeneous degree distribution of the form $p(k) \propto k^{-\alpha}$ where $\alpha > 0$. These networks are termed "scale-free," and are marked by the presence of many low degree nodes and a few "hubs" – highly connected nodes which, for example, could be celebrities on Twitter.

Barabási and Albert proposed a random graph model, with only one parameter m, which grows networks that have a scale-free degree distribution ($p(k) \propto k^{-3}$). The model starts with m nodes and no edges and repeatedly adds nodes. Each new

node forms m links according to preferential attachment, leading to a rich-get-richer process.

Clustering and the small-world phenomenon Another important property, especially for social networks, is clustering. A highly clustered network is a network in which "a friend of your friend is also your friend." That is, the network is populated by many triangles. The clustering coefficient is a measure of the number of triangles in a network. There exist several definitions, but the one we use is the transitivity $C_\triangle$ given by

$$C_\triangle \equiv \frac{\#\text{ closed triples}}{\#\text{ triples}} = \frac{3\times\ \#\text{ triangles}}{\#\text{ triples}}.$$

Similar to edge density, $C_\triangle$ is a number in $[0, 1]$ which counts the fraction of all possible triangles which are observed in the network.

Real-world social networks have high clustering, but at the same time a low diameter (maximum distance between nodes) or average shortest path length (the average distance between two nodes). This combination is atypical – generally high clustering indicates that the network is structured (i.e. a lattice), whereas low diameter suggests the network is not structured (i.e. an ER network). Watts & Strogatz proposed a random graph model which generates networks that can exhibit this property [17], which they termed the small-world property. The model starts with a lattice, and then rewires each link with probability p. Increasing p monotonically decreases the expected clustering and diameter, but for small (but nonzero) values of p, the resulting network has both high clustering and low diameter.

Community structure and modularity People can naturally be classified based on group-belongings: political affiliation, social class, hobbies, etc. In a network,

a community is loosely defined as a group of nodes which has more within-group connections than connections to other groups. A network is said to have community structure/be modular if it has communities which are distinctly separate from each other. The most common way to quantify the existence of community structure is by defining a function $f(G, \mathcal{P})$ on the network, with $\mathcal{P}$ a partition of the network into groups. This function f, when maximized over all $\mathcal{P}$, should be larger for modular networks. A popular function is modularity, Q [18]. Suppose we have a partition $\mathcal{P} = \{c_i\}_{i=1}^N$ where c_i is the community which node i belongs to. Modularity is defined as

$$Q \equiv \frac{1}{2M} \sum_{(i,j) \in V \times V} \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(c_i, c_j)$$

where $A = [a_{ij}]$ is the *adjacency matrix* of the network ($a_{ij} = 1$ if $(i, j) \in E$, $a_{ij} = 0$ if $(i, j) \notin E$), k_i is the degree of node i, and δ is the Kronecker delta, The optimal partition can found using a community detection algorithm [19].

The stochastic block model (SBM) is the prototypical random graph model for creating networks with community structure [20, 21]. The model is specified by the number of nodes N, the number of communities r, a community labelling $\{c_i\}_{i=1}^N \in \{1, \ldots, r\}^N$ of the nodes, and P which is a $r \times r$ matrix such that P_{ij} gives the probability those a node in community i connects to a node in community j. For every pair $(u, v) \in V \times V$, with probability $P_{c_u c_v}$ the edge (u, v) is present.

Assortative mixing Another property of social networks is that similar people tend to link with similar people. Generally, similar means "similar degree," and this phenomena is termed homophily in the social science literature or assortative mixing in the network science literature [22]. If other information (metadata) is known about

the nodes, then that may also explain the network structure. The degree assortativity coefficient $\gamma \in [-1, 1]$ is defined as the Pearson correlation of the degree of endpoints u, v over all links (u, v). When γ is negative, the network is said to be disassorative – low degree nodes tend to link to high degree nodes. When γ is positive the network is said to be assortative – high degree nodes link with high degree nodes, low degree link with low degree. Social networks tend to be assortative whereas technological networks are often disassortative [22].

2.2 Network dynamics

Here we define some of the traditional models of information flow. We divide this section into three parts. In Sec. 2.2.1 we treat models coming from, or inspired by, the study of biological contagion. In Sec. 2.2.2 we study the dependence of these models on network structure and highlight the interesting features of complex contagion. In Sec. 2.2.3 we introduce models from the field of "opinion dynamics." For comprehensive reviews, see [23, 24] on models from epidemiology and [25, 26] on models from opinion dynamics.

2.2.1 Epidemiology and information flow

Simple contagion In epidemiology, the most commonly used models for predicting and understanding disease spread belong to a class called compartmental models [23]. These models divide a population into compartments based on the state of each individual, and within each compartment all individuals are considered identical. The first compartmental models were proposed in 1927 by Kermack and Mckendrick [27]

and in 1928 by Reed and Frost, though not published until the 1950s [28]. In the SI model, possible states include (S)usceptible and (I)nfected. Susceptible individuals move to the infected state after being exposed to an infected individual, which occurs with some constant rate α. In the SIR model, a (R)ecovered state is included, in which individuals are effectively removed from the population. Infected individuals move to the recovered state with some constant rate β. An underlying social network constrains interactions between individuals. There are numerous variants of these models, such as the SIS model in which infected individuals can become susceptible again.

Complex contagion Social (or "complex") contagion are better explained by threshold models [29, 30]. Like the SI model, there are two states: susceptible and infected. However, the probability for a node u to become "infected" is a nonlinear function of the number of neighbors which are already infected. A threshold ϕ is specified at the beginning of a simulation, and the probability of a susceptible node u becoming infected is nonzero only if the fraction of infected neighbors of u is greater than ϕ. A stochastic version of this model allows for specifying a distribution of thresholds. In a social context, this thresholding process is intended to mimic social reinforcement – that is, the behavior of an individual is determined by the fraction of their neighbors which they see already engaging in that behavior. This seemingly innocuous mechanism is critical for explaining how information and ideas spread in social networks. [31, 32]. An interesting model introduced by Dodds & Watts [33] was shown to interpolate between simple contagion models and threshold models.

Information cascades Another similar model is the Independent Cascades (IC) model for information diffusion. The dynamics of the IC model are in fact equivalent to the famous bond percolation model, and the final state can be mapped to that of an SIR model [34, 35]. The IC model is well-studied in the computer science community for its theoretical "influence maximization" problem [4, 36]. As with the previous models, the IC model involves starting with an initial seed of infected nodes. At each discrete time step t, every newly activated node u gets one chance to infect each uninfected neighbor v, and succeeds with probability p_{uv}, a parameter specified at the start of the simulation. If u infects v then in time step $t+1$, node v will get a chance to infect its neighbors. This process continues until no more spreading is possible. The resulting collection of edges traced by the infection forms a tree (because if u infects v and v infects w, it is impossible that w infects u). In [36] the authors showed how the IC model could be extended to complex contagion. Kempe et al. applied the IC model to modeling the propagation of recommendations [36], Gomez-Rodriguez et al. used the IC model to study how information flows through blogs [37], and Goldenberg et al. studied the dependence of the IC model on network structure [38].

2.2.2 Features of complex contagion

The simple contagion models used to describe biological contagion and complex contagion models used to describe social contagion contrast greatly in their resulting dynamics. Namely, Watts proposed a model for complex contagion and showed there exists a cascade window where the final outbreak size is non-monotonically related to the network density [30]. That is, increased network density *can* inhibit diffusion whereas in simple contagion increased network density always promotes diffusion. It

should be highlighted that this feature depends on the threshold and is not observed in all complex contagion models [39]. Closely related to network density, clustering plays an important and distinct role in the spreading of contagion. In simple contagion, it is well-known that clustering inhibits diffusion [40]; however, the role of clustering in the spreading of complex contagion is less clear. In fact, studies show that clustering can either promote or inhibit contagion [23, 39, 41]. Finally, in complex contagion, long ties (edges which bridge communities) can inhibit diffusion [42], giving rise to the saying "weakness of long ties." This entirely opposes the "strength of weak ties" result [43] that shows long ties are extremely important to the efficient spreading of simple contagion.

2.2.3 Social and opinion dynamics

There is a breadth of literature in the field of social/opinion dynamics, which also attempts to model information flow. Here the state of a node is its opinion, and at each time step a node(s) updates their opinion. The first goal is to understand if the system will reach consensus on an opinion or if there is coexistence. If there is consensus, the second goal is to determine how quickly the system reaches consensus.

DeGroot learning The DeGroot model [44] assumes that each individual i has a continuous-valued opinion $x_i(t) \in [0, 1]$ which is updated in time step $t + 1$ by averaging all of i's neighbors' opinions at time t. This can compactly be represented by the equation

$$\mathbf{x}(t + 1) = A\mathbf{x}(t)$$

where A is the adjacency matrix of the network (normalized by dividing each row by k_i so that each row sums to unity). Then A is a stochastic matrix and the above equation describes a Markov chain. Familiar conditions for the convergence of Markov chains [45] give corresponding conditions for when there is consensus on an opinion $\mathbf{x}(\infty) \in [0,1]^N$.

Voter model Unlike the DeGroot model, the traditional "voter model" [46, 47] supposes each individual i has a discrete opinion $x_i \in \{0,1\}$. For example 0/1 may represent the democratic/republican candidates. The voter model is a Markov process which can be described in discrete or continuous time. In the discrete time version, at time $t+1$ an individual i is chosen uniformly at random to be updated. One of their neighbors j is chosen uniformly at random, and $x_i(t+1) \leftarrow x_j(t)$. The dynamics stop after T time steps and the quantity of interest is the average density of 1's,

$$\rho(T) \equiv \frac{1}{N} \sum_{i=1}^{N} x_i(T).$$

An unrealistic feature of the voter model is that on any finite network ρ will reach 0 or 1 starting from any initial configuration. That is, everyone eventually shares the same opinion. To deal with this unrealistic issue, several variants of the voter model have been proposed in which consensus is either slowed or prevented altogether. Some variants include: noise (random changes in opinion), multiple states, zealots & contrarians (individuals that never change their opinion or that tend to go against their neighbors), aging/inertia (the longer an individual has had an opinion, the lower their chance of copying a neighbor), nonlinear rates (the chance of copying is a nonlinear function of the fraction of neighbors holding opposing opinions), and

concealed & public opinions. See [48] for a survey of voter models.

Axelrod model The Axelrod model [49] was introduced as a model for the dissemination of culture. We cover it here because it incorporates a desirable property, social homophily, and gives rise to interesting dynamics which could be applicable to information flow. It is based on two principles: (i) individuals become more similar when they interact, (ii) individuals interact more often with people that are similar to them. The model assumes each node u has F cultural traits $\sigma_f(u)$ where $f = 1, \ldots, F$. Each trait can take on values in $\{1, \ldots, q\}$ and thus a node's state is represented by a vector $\sigma(u) \in \{1, \ldots, q\}^F$. In a single time step, a node u and a neighbor v are selected at random. The average overlap between their states is computed as

$$\omega(\sigma(u), \sigma(v)) \equiv \frac{1}{F} \sum_{f=1}^{F} \delta(\sigma_f(u), \sigma_f(v)).$$

With probability $\omega(\sigma(u), \sigma(v))$, nodes u and v interact and one of the traits f which is different is selected at random. The update puts $\sigma_f(v) \leftarrow \sigma_f(u)$, i.e. the trait spreads to the neighbor. An interesting feature of this model is that the dynamics depend heavily on the number of options for each trait, q. When q is small, individuals initially share many traits and so consensus is quickly reached. When q is large, interactions are rare and fragmented cultural domains emerge.

2.3 Information theory and entropy

The goal of this section is to introduce the field of information theory and some of the relevant information-theoretic measures, concluding with the estimator which

will be our focus. Throughout this section, we use capital letters to denote random variables. Specifically, we are interested when X is a discrete random variable with finite support $\mathcal{X}$ (also called the alphabet of X) and probability mass function (PMF) denoted $p(\cdot)$. For this section, we assume all logarithms are base-2.

2.3.1 Information theory

Information theory is a discipline that was born out of Claude Shannon's 1948 paper *A Mathematical Theory of Communication* [50]. The main goal of the discipline is to rigorously prove limits of data compression and communication. Consider the following setting (this example has been adopted from [51]): a source produces a sequence of symbols $X_1, X_2, \ldots$, which are drawn according to some distribution X with support $\mathcal{X}$. For now suppose $\mathcal{X} = \{a, b, c\}$ and $p(a) = 0.7$, $p(b) = p(c) = 0.15$. We wish to encode the string $X_1, X_2, \ldots$ in binary. That is, we want to devise a function $C(x)$ which maps from $\mathcal{X}$ to finite length strings of $\{0, 1\}$. The function should satisfy the prefix condition, that is the codeword $C(x_1)$ does not appear at the start of any other codeword $C(x_2)$ (i.e. $C(x_1) = 0$ and $C(x_2) = 01$ break the condition). Since there are only 3 symbols, a possibility is to use 2 bits for each symbol. We could have $C(a) = 00$, $C(b) = 01$, and $C(c) = 10$. The expected length of a codeword is 2 bits per symbol. We can improve this by exploiting the fact that symbol a occurs very often. Hence we could give a single bit to a, i.e. $C(a) = 0$, and then set $C(b) = 10$, $C(c) = 11$. Now the expected code length is only $0.7 \times 1 + 0.15 \times 2 + 0.15 \times 2 = 1.3$ bits per symbol. It turns out we can improve this further by encoding two symbols at once and assigning $C(aa) = 0$, since $p(aa) = 0.7^2 = 0.49$ means that half the sequence will be aa. Encoding other pairs

according to the (optimal) Huffman encoding scheme results in a average code length of 1.20. Shannon proved that the optimal code length is equal to the entropy of the source $H(X)$ which will be defined in the next section. Encoding triples and so on via Huffman encoding can produce average code lengths which are arbitrarily close to the entropy, although this will increase compression time.

2.3.2 Entropy

The entropy of a discrete random variable X is a measure of the uncertainty in X. It can be derived either as we saw above or as the only functional which satisfies four desirable properties of information [50]; however, we will not detail this approach here. The entropy is defined as the expected surprise, $-\log p(X)$.

$$H(X) \equiv \mathbb{E}_p[-\log p(X)]. = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

In the example from the last section,

$$H(X) = -0.7 \log(0.7) - 0.15 \log(0.15) - 0.15 \log(0.15) = 1.18$$

which was nearly achieved by the pair encoding, having an average code length of 1.20. There is a long list of interesting mathematical properties of $H(X)$. We only note that $H(X) \geq 0$ for any random variable X, and that for a fixed support $\mathcal{X}$, $H(X)$ is maximized when X is the uniform distribution on $\mathcal{X}$ (with $H(X) = \log |\mathcal{X}|$) and minimized when X is a point mass (with $H(X) = 0$).

Commonly we wish to quantify the dissimilarity between two probability distri-

butions. Suppose X and Y are random variables with PMFs p and q, respectively, defined on the same support $\mathcal{X}$ (we can relax this by assuming only that X is absolutely continuous with respect to Y, i.e. $q(x) = 0$ implies $p(x) = 0$ so that division by zero never occurs). The Kullback-Leibler (KL) divergence is an asymmetric measure which is a natural extension of the Shannon entropy. It is defined as

$$D_{KL}(p \mid\mid q) \equiv -\sum_{x \in \mathcal{X}} p(x) \log \frac{q(x)}{p(x)} = -\mathbb{E}_p[\log q(X)] + \mathbb{E}_p[\log p(X)] = H(p, q) - H(X)$$

where $H(p, q)$ is the cross-entropy between p and q defined as

$$H(p, q) \equiv H(X) + D_{KL}(p \mid\mid q) = -\mathbb{E}_p[\log q(X)] = -\sum_{x \in \mathcal{X}} p(x) \log q(x).$$

The cross-entropy is the expected surprise in Y, where the expectation is incorrectly taken according to X's distribution. This quantifies the expected number of bits to encode Y, given that we incorrectly believe Y follows $X's$ distribution; while the KL divergence is the additional number of bits. It is easy to see that $H(X, Y) \geq H(X)$.

Other useful quantities are the joint entropy, conditional entropy, and mutual information. The joint entropy of X, Y is simply the expected surprise in the joint random variable (X, Y) and is defined as

$$H(X, Y) \equiv -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log p(x, y).$$

The conditional entropy of X given Y is

$$H(X|Y) \equiv -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log p(x|y).$$

The mutual information in X and Y is defined as

$$I(X;Y) \equiv -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}.$$

Notice if X and Y are independent then the argument of the logarithm is 1 and so $I(X;Y) = 0$. In fact, $I(X;Y) = 0$ if and only if X and Y are independent. The mutual information thus measures the independence of X and Y. Another useful way to write $I(X;Y)$ is

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Suppose now that we have a sequence of random variables $X_1, X_2, \ldots$ (not necessarily identically distributed) and we want to quantify the long-time uncertainty. In the context of dynamical systems, X_n may be the predicted temperature on day n and we want to quantify our uncertainty in the weather forecasts. It does not seem appropriate to use our uncertainty in predicting the temperature today or tomorrow as an estimate of the "uncertainty in the weather," rather we should use a average over time. The information rate is defined as the limit of the joint entropy per symbol

$$h(\{X_n\}) \equiv \lim_{n\to\infty}\frac{1}{n}H(X_1,\ldots,X_n).$$

For stationary processes this reduces to

$$h(\{X_n\}) = \lim_{n\to\infty} H(X_n|X_1,\ldots,X_{n-1}).$$

The information rate is a key concept in the Asymptotic Equipartition Property [52];

allowing us to classify sequences as either typical or atypical.

Entropy and related quantities have many applications to dynamical systems. For two stochastic processes X and Y, the transfer entropy $T_{X\to Y}$ describes the uncertainty in predicting the future of Y given the past of X [53]. It is a non-linear, model-free generalization of Granger causality [54, 55]. Transfer entropy has been applied in neuroscience, geoscience, and finance. An extension, the causation entropy $C_{X\to Y|Z}$, is the transfer entropy conditioned on a third process Z [56, 57]. Unlike the transfer entropy, the causation entropy allows for determining indirect influence.

One of the challenges of information-theoretic approaches is in efficiently estimating the measures set forth. Generally a significant amount of data is required for the estimators to converge. This is of concern in neuroscience where data is expensive, although not so concerning for the application to social media. Much research has been done on improving the computational and data efficiency of information-theoretic estimators such as conditional mutual information, cross-entropy, and transfer entropy. For example, symbolic transfer entropy (STE) has been developed as an efficient alternative to transfer entropy [58].

We now present the estimator which will be the focus of the following work. For context, suppose we have text data which was generated on an online social media platform such as Twitter or Facebook. Each word can be encoded as an integer, and we can interpret the text as a random process indexed by the position of each word. We wish to calculate the entropy rate h of this process. It is challenging to estimate the entropy rate h for natural language data because information is present in both the ordering of the words and the relative frequencies of words [59]. To this end,

Kontoyianni et al. [60] proved that the estimator

$$\hat{h} \equiv \frac{T \log T}{\sum_{t=1}^{T} \Lambda_t}, \tag{2.1}$$

converges to the true entropy rate h of the text, where T is the length of the text and Λ_t is the match length of the prefix at position t: it is the length of the shortest string of words starting at t that has not previously appeared in the text. Alternatively, it is one more than the length of the longest string of words starting at t that appears in the past text.

2.4 Measuring information flow: the cross-entropy

Equation (2.1) generalizes to an estimator of the **cross-entropy** $h_\times$ between two texts A and B [61, 62]:

$$\hat{h}_\times(A \mid B) \equiv \frac{T_A \log T_B}{\sum_{t=1}^{T_A} \Lambda_t(A \mid B)}, \tag{2.2}$$

where T_A and T_B are the lengths of the two texts, and $\Lambda_t(A \mid B)$ is the length of the shortest substring $[A_t, A_{t+1}, \ldots, A_{t+\Lambda_t(A|B)+1}]$ starting at position t of text A not previously seen in text B. Previously, in this case, refers to all the words of B written prior to the time when the tth word of A was written. Specifically, we compute $\Lambda_t(A \mid B)$ by searching for each substring $[A_t]$, $[A_t, A_{t+1}]$, ... within $B_{:t} \equiv [B_j \mid \text{time}(B_j) < \text{time}(A_t)]$, the ordered sequence of words in B that appear before

the time of the t-th word in A, until the first substring $[A_t, \ldots, A_{t+\Lambda_t(A|B)+1}]$ that is not seen in $B_{:t}$. By matching the future text of A (words posted at times $\geq \text{time}(A_t)$) against the past text of B (words posted at times $< \text{time}(A_t)$) at every t, only B's past predictive information about A's future is estimated and temporal precedence is satisfied. We can then measure the information flow from individual B to individual A by computing the cross-entropy $\hat{h}_\times(A \mid B)$ using their two text streams. Low values of $\hat{h}_\times(A \mid B)$ indicate that B is very predictive of A, which we associate with information flow. Higher $\hat{h}_\times(A \mid B)$ indicates B is not predictive of A, which suggests low information flow from B to A. We briefly note that this estimator can be generalized beyond the pairwise setting, just as causation entropy generalizes transfer entropy. Implementation details are described in [62].

We introduce one last information-theretic measure which is functionally equivalent to the entropy (or cross-entropy). The predictability, Π, is obtained from Fano's Inequality [52], and offers an upper bound on how accurately an ideal predictive algorithm can perform when working with data of a given entropy. The predictability is the probability that such an algorithm will correctly predict the next word.

$$h(\Pi) + (1 - \Pi)\log(z - 1) \geq h_\times \tag{2.3}$$

where $h(\Pi) \equiv -\Pi\log(\Pi) - (1 - \Pi)\log(1 - \Pi)$ and z is the cardinality of the sample space; in our problem, this is the vocabulary size or number of unique words for the quoter model (Sec. 3.1.1). The predictability is then given by finding numerically the largest Π that satisfies Eq. (2.3). Equation (2.3) demonstrates that $h_\times$ and Π are functionally equivalent (and inversely related, with higher $h_\times$ corresponding to lower Π and vice versa) as z is a constant for the model we study here. Higher values of Π

(lower $h_\times$) correspond to higher amounts of information flow.

2.5 Modeling information flow: the quoter model

We now introduce the quoter model, which was first proposed by Bagrow and Mitchell in [1]. The quoter model is an idealized model for how people communicate and generate text through online social media platforms such as Twitter or Facebook, although it may afford itself to more general communication. The idea behind the quoter model is that people tend to copy (quote) other people. This may occur for one of two reasons: (i) the individual is reposting/retweeting someone else (ii) the individual is replying to someone else and is borrowing a short phrase. For an example of the latter, if Bob asks Alice "would you like to get lunch tomorrow," Alice may reply "I would love to get lunch." The words "to get lunch" are quoted, and "love" is roughly a rewording of "like." The quoter model takes a social network and generates text streams for each individual by a combination of sampling from a given vocabulary distribution or by quoting words from their neighbors. The dynamics of the model are dependent on the quote probability, q, which determines the frequency of quoting as opposed to random sampling. In the case $q = 0$, the text streams are completely random and there will be no information flow, and no dependence on network structure. Implementation details are discussed later in Sec. 3.1.1. The result of simulating the quoter model for T time steps, is a collection of N text streams, which we can apply our pairwise measure of information flow – the cross-entropy.

2.6 Related work

An information-theoretic approach was used to study the dynamics of human coordination in [63]. The authors used symbolic transfer entropy (STE) to quantify temporal correlations between Twitter users. The collection of tweets consisted of posts related to "collective events" such as those containing a hashtag pertaining to a protest. The value of STE notably depends on a temporal resolution parameter, and the authors found that these events were marked by a change in "temporal scale". That is, they found that the maximum STE was found at a different temporal resolution. Closely related, Ver Steeg and Galstyan (2012) used transfer entropy to study influence on Twitter, and consequently infer social networks [64]. We crucially mention that both of these studies employ information-theoretic measures only to the timings of tweets, and do not examine the textual content of the tweets. Ver Steeg and Galstyan (2013) extended their previous approach to involve topic modeling [65]. They used transfer entropy on the timings and the topics extracted from the text to quantify influence. De at al. [66] proposed a model for opinion dynamics in which individuals hold a latent continuous opinion and post new messages based on the entire history of their neighbor's messages before updating their opinion. This model is similar to the quoter model and the DeGroot model. Although the dynamics depend on the entire history of the process, the quantity of interest is only the final opinion. Finally, the cross-entropy estimator has been used to study the limits of predictability of mobility patterns and social media posting. [62, 67].

Chapter 3

The quoter model and complex contagion

In this chapter, we simulate the quoter model on various synthetic and real-world networks to assess the impact of network structure on information flow. We make comparisons with simple and complex contagion models. This study was published in [68].

3.1 Materials and methods

Here we describe the methods used to simulate the quoter model and measure information flow over a network. We describe the network features we study in relation to information flow, and we provide the details on the synthetic (random graph models) and real-world network datasets we study.

3.1.1 Simulating the quoter model

The following section describes the parameters of the quoter model and the process used to simulate the quoter model. The quoter model requires a network $G = (V, E)$ with $|V| = N$ nodes and $|E| = M$ edges. In our simulations, at each time step t a node v (also called the *ego*) is randomly selected. A message length m is chosen randomly according to a Poisson distribution with mean λ. Node v generates text (encoded as an integer in $\{1, \ldots, z\}$) according to one of two mechanisms:

- *Quoting a neighbor.* With probability q, node v picks a random neighbor u (also called the *alter*) to quote from. A contiguous block of m words is randomly chosen from node u's text up to and including time $t - 1$.

- *Randomly generating new content.* With probability $1 - q$, node v generates m words at random according to their vocabulary distribution. We assume the vocabulary distribution follows a Zipf law which is commonly observed in real-world language usage patterns [1]. The Zipf law says that the probability of observing word w with rank $r_w \in \{1, \ldots, z\}$ is given by $W(w) = H_{z,\alpha}^{-1} r_w^{-\alpha}$ where $\alpha > 0$ is the Zipf exponent and $H_{z,\alpha} = \sum_{r=1}^{z} r^{-\alpha}$ is the generalized harmonic number.

In either case, the m words are appended to node v's text and marked with a time stamp t (these time stamps are only important when computing $h_\times$). Time then increments by one, and the process repeats until $T = 1000N$ time steps have elapsed so that each node has created a text stream of approximately $1000\lambda = 3000$ words. The choice of $T = 1000N$ was made so that the entropy estimator would converge (see [60, 61] for convergence proofs). For our simulations, unless otherwise specified,

we assume $q = 1/2$, $\lambda = 3$, $z = 1000$, $\alpha = 1.5$ as used in [1]. We investigate other parameter choices in Sec. 3.2.4 and App. A.1.

3.1.2 Measuring information flow over the network

The result of simulating the quoter model is N text streams approximately of length 3000. We then compute and record $h_\times(v|u)$ for all $(u, v) \in E$, according to the cross-entropy estimator (Eq. 2.2). We perform several simulations for a fixed network and then compute $\langle h_\times \rangle$ and $\mathrm{Var}(h_\times)$, the mean and variance of $h_\times$ respectively. The mean and variance are thus computed over several edges for a single simulation, and over several simulations. In App. A.2 we investigate the distribution of $h_\times$ and give support for using the mean and variance as appropriate summary statistics. Since $h_\times$ is negatively correlated with information flow (higher $h_\times$ implies lower information flow) we also compute the predictability Π, which is obtained from Fano's Inequality [52], and is functionally equivalent to $h_\times$ (since the vocabulary distribution is the same for all nodes), but is positively correlated with information flow. This quantity allows us to directly compare our view of information flow with traditional contagion approaches.

3.1.3 Simulating contagion models

To compare and contrast information flow in the quoter model, we also simulate traditional models of information flow, specifically simple and complex contagion. For simple contagion we simulate a stochastic SIR model on different networks (1000-

node Erdős-Rényi and Barabási-Albert networks, as well as a sample of real-world networks) using [69]. For the simulations here we set the transmission rate 20 and recovery rate 1. We initialize with a random 5% of the nodes infected, and run 10 outbreaks on 100 realisations of the network for each choice of average degree $\langle k \rangle$. For complex contagion we use exactly the same parameters, except we introduce a threshold function for transmission as in [30], where the transmission rate is set to zero if the proportion of infected neighbors is below some threshold ϕ (and we set $\phi = 0.18$ following [30]). For all simple and complex contagion simulations we measure the peak outbreak size, noting that larger outbreak sizes conventionally correspond to greater information flow.

3.1.4 Assessing the impact of structure on dynamics

In this work we use a number of network models (random graphs) tailored to control for various network properties such as density, clustering, and modular structure. Here we describe the models and properties we study in relation to information flow in the quoter model.

Density and average degree To explore how network density relates to information flow, we create Erdős-Rényi and Barabási-Albert networks of N nodes with varying average degree, $\langle k \rangle$, allowing us to the tune their densities. For the Erdős-Rényi networks we add edges independently with probability $p = \langle k \rangle / (N-1)$. For the Barabási-Albert model we start with $m = \langle k \rangle / 2$ nodes with no edges and add nodes which each form m links with previous nodes according to preferential attachment.

Here we measure how cross-entropies varies with the densities of the networks using their average degree $\langle k \rangle$ and edge density $M/\binom{N}{2}$ where M is the total number of edges in the network. To complement the Erdős-Rényi and Barabási-Albert results, we also compare the densities of real networks with their average cross-entropy.

Degree heterogeneity To assess the role of degree heterogeneity on information flow, we study the simplest random graph model with tunable degree heterogeneity, termed "dichotomous networks" in [70, 71]. Dichotomous networks are generated via the configuration model. They have only two types of nodes – those with degree k_1 and those with degree k_2. We assume there are $N/2$ nodes of each degree and fix $k_1 + k_2$ so that the average degree is fixed. The mean and variance of the degree distribution, respectively, are given by $\mu = \frac{1}{2}(k_1 + k_2)$ and $\sigma^2 = (k_1 - k_2)^2/4$. We are interested in how the cross-entropy varies with k_1/k_2. When $k_1/k_2 = 1$ the network reduces to a random k-regular graph ($\sigma^2 = 0$), while $\sigma^2 \to \infty$ as $k_1/k_2 \to 0$.

Clustering We studied the role of clustering using two methods. First, we constructed "small-world" networks using the Watts-Strogatz (WS) model described in Sec. 2.1.2. We varied the rewiring probability p to generate networks with varying clustering and diameter.

As mentioned in Sec. 2.2.2, previous research has showed mixed results when studying the impact of clustering on spreading. These mixed results are due to the challenge of generating networks with tunable clustering but for which other properties, such as density or diameter, can be controlled for. We thus should be careful in drawing conclusions from one demonstration. With this skepticism, we turn to another method: we apply the established stochastic rewiring or "x-swap"

method [72–74] to real-world networks. The x-swap method repeatedly chooses two links at random and two randomly selected endpoints of those links are swapped (provided that the number of links does not change by swapping and the network does not become disconnected). The resulting network has lower clustering but the degree distribution is preserved. We performed $5M$ swaps for each real network, where M is the number of edges, and then compared $h_\times$ for the original and x-swapped networks to assess the role of clustering on information flow.

Community structure and modularity To investigate the importance of community structure in information flow, we generated networks according to the stochastic block model, explained in Sec. 2.1.2. Each network has N nodes and two planted blocks of nodes, denoted A and B, of equal size $m \equiv N/2$. Here there are two connection probabilities: p_0 (the within-block connection probability) and p_1 (the between-block connection probability) governing the probability for a link to form between nodes in the same block and in different blocks, respectively. The expected modularity in this two-block stochastic block model is

$$Q = \frac{1}{2}\left(\frac{p_0 - p_0 m + p_1 m}{p_0 - p_0 m - p_1 m}\right).$$

Our main quantities of interest are the average cross-entropy on within-block edges, $\langle h_\times(\text{within})\rangle$, the average cross-entropy on between-block edges $\langle h_\times(\text{between})\rangle$ and their difference, $\Delta h_\times \equiv \langle h_\times(\text{between})\rangle - \langle h_\times(\text{within})\rangle$. These quantities describe to what extent information flows within and between communities.

We also computed modularity for real networks using the Louvain method [19]. The Louvain method is a hierarchical community detection algorithm that finds a

partition of nodes that maximizes modularity. As commonly done, we initialize each node in its own community.

A heterogeneous population Previous research [75, 76] has shown that heterogeneity in the dynamical parameters (such as recovery rates in the SIR model) can be as important as structural heterogeneity. Communities offer an obvious way to implement such heterogeneity. We again use the two-block SBM, but now distinguish the two groups A and B by giving them different Zipf exponents α_A, α_B, respectively, for their vocabulary distributions. A larger α (steeper distribution) corresponds to a less diverse vocabulary, and could capture a group of people that is more consistent and repetitive in their dialogue. In contrast, a lower α (shallower distribution) may describe a group of people that use more diverse words.

3.1.5 Network datasets

To supplement the above graph models, we also studied contagion and quoter model dynamics on real-world networks. We developed a corpus of 10 social networks spanning a range of sizes and densities that were used as the basis for simulation. See App. B for details on network sources and processing. Table 3.1 shows several descriptive statistics for the networks we analyzed.

3.2 Results

Here we compare information flow in the quoter model with traditional simple and complex contagion (Sec. 3.2.1), then investigate how degree heterogeneity (Sec. 3.2.1),

Table 3.1: Descriptive statistics for real-world networks used in this study. ASPL: Average Shortest Path Length. Modularity computed using the Louvain method [19].

Network	$\|V\|$	$\|E\|$	$\langle k \rangle$	Density	Transitivity	ASPL	Modularity	Assortativity
Sampson's monastery	18	71	7.9	0.464	0.53	1.54	0.29	−0.07
Freeman's EIES	34	415	24.4	0.740	0.82	1.26	0.07	−0.15
Kapferer tailor	39	158	8.1	0.213	0.39	2.04	0.32	−0.18
Hollywood music	39	219	11.2	0.296	0.56	1.86	0.20	−0.08
Golden Age	55	564	20.5	0.380	0.53	1.64	0.45	−0.13
Dolphins	62	159	5.1	0.084	0.31	3.36	0.52	−0.04
Terrorist	62	152	4.9	0.080	0.36	2.95	0.52	−0.08
Les Miserables	77	254	6.6	0.087	0.50	2.64	0.56	−0.17
CKM physicians	110	193	3.5	0.032	0.16	4.24	0.61	0.11
Email Spain	1133	5452	9.6	0.009	0.17	3.61	0.57	0.08

clustering (Sec. 3.2.2) and network modularity (Sec. 3.2.3) affect information flow. We also study how heterogeneity in the parameters affects information flow compared to the effects of network structure (Sec. 3.2.4).

3.2.1 Information flow and models of contagion

In Sec. 2.2.2 we reviewed the distinguishing features of simple and complex contagion. We know increased density is associated with increased spreading in simple contagion, while complex contagion exhibits a non-monotonic relationship with density. Figures 3.1A,B illustrate this difference. For the traditional contagion models we use the average peak size of the outbreak to measure information flow. For the quoter model we use the average predictability, Π. It should be noted that it is unclear what the mapping between average peak size and predictability is. Thus we limit our conclusions to qualitative statements about how the trends compare. From Fig. 3.1A we see that density promotes information flow in simple contagion. Figure 3.1B confirms a non-monotonic trend with density; density promotes spreading up to a certain point (indicated by the dashed line), and then subsequently shows an opposite effect,

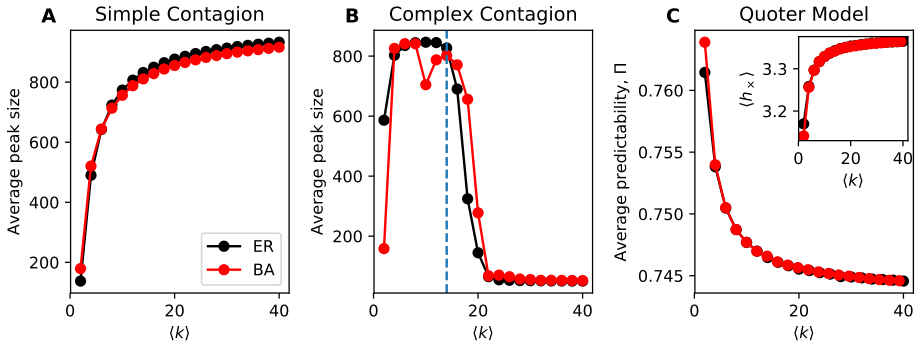

*Figure 3.1: Denser networks are associated with higher information flow for simple contagion but lower information flow for both complex contagion and the quoter model. Here density is measured by average degree $\langle k \rangle$ for Erdős-Rényi (ER) & Barabási-Albert (BA) model networks. (**A**) Simple contagion: SIR model. (**B**) Complex contagion: SIR model with thresholding mechanism as in [30], with threshold $\phi = 0.18$. (**C**) Quoter model. (Panel C, inset) Average cross-entropy on links; higher cross-entropies correspond to lower predictabilities and lower information flow, unlike for contagions where higher average peak sizes correspond to higher information flow. Networks consisted of $N = 1000$ nodes and each point constitutes 200 simulations; parameters for simulating information flow in these models are described in full detail in Sec. 3.1.*

which is mimicked by the quoter model in Fig. 3.1C. These results are supported by simulations on real-world networks shown in Fig. 3.2.

In Fig. 3.1C we see that Erdős-Rényi (ER) and Barabási-Albert (BA) networks are qualitatively indistinguishable in terms of information flow. It is rarely the case that dynamical processes on ER and BA networks show similar outcomes. Hubs generally play an important role in mediating diffusion. We investigate this further by examining the variance of $h_\times$ in ER and BA networks in Fig. 3.3A. We observe that $h_\times$ is slightly more variable in BA networks than ER networks, and that this difference is exaggerated for smaller networks.

To further explore the role of degree heterogeneity, we investigate dichotomous networks (Sec. 3.1.4). Here half the nodes have degree k_1 and the other half have degree k_2. We vary the degree ratio k_1/k_2 to tune the variance of the degree distri-

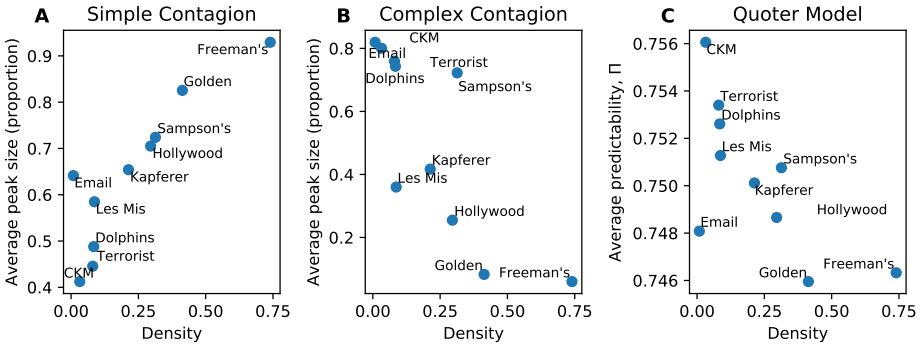

*Figure 3.2: Information flow on real-world networks. (**A**) Simple contagion: SIR model. (**B**) Complex contagion: SIR model with thresholding mechanism as in [30], with threshold $\phi = 0.18$. (**C**) Quoter model. Here information flow measures (average peak size, average text predictability) are compared to network density $M/\binom{N}{2}$. The association between information flow and density, either positive (simple contagion) or negative (complex contagion, quoter model), is significant (Wald test on non-zero regression slope, $p < 0.05$). Each point constitutes 300 simulations. Parameters for simulating information flow in these models are described in full detail in Sec. 3.1.*

bution. In Fig. 3.3B we observe that degree heterogeneity has little impact on the average cross-entropy, as we saw in Fig. 3.1C. The average degree plays the largest role in mediating information flow. A minor observation is that smaller networks seem to have slightly higher information flow. Figure 3.3C again shows that degree heterogeneity relates to the variance of the cross-entropy. Heterogeneous networks have more heterogeneous $h_\times$. That is, there exists overly influential (or overly predictable nodes). Which nodes are overly predictable? We answer this by computing conditional expectations of $h_\times$ conditioned on degree of the ego (the node being predicted) and alter (the node predicting). In Fig. 3.3D we see that the degree of the ego but not the alter plays a role in the information flow: degree-k_1 egos have more information flow than degree-k_2 egos regardless of the degree of the alter.

3.2.2 Interplay of clustering and information flow

We now study how clustering (measured according to 2.1.2) affects information flow. Clustering plays a complicated role in both simple and complex contagion [39,40] and we report interesting, if mixed, results in Fig. 3.4 with the quoter model's information flow.

In Fig. 3.4A we create Watts-Strogatz networks with tunable clustering. From the top panel, we see that lower rewiring, and thus higher clustering, is associated with increased information flow. This result is robust to different choices for the average degree and network size. Network size again plays a negligible role. The bottom panel shows how the rewiring probability changes the clustering and average shortest path length (ASPL) simultaneously. Matching the top and bottom panels, we see the largest change in information flow occurs when the clustering starts to drop but the ASPL is relatively constant. This observation supports that clustering, and not ASPL, is the cause of increased information flow.

In Fig. 3.4B we show the result of applying the x-swap randomization procedure to real-world networks. We compute $h_\times$ on both the original and x-swapped networks. The x-swapped networks have lower clustering than the original. Half of the networks show a decrease in $h_\times$ while half show an increase, weakening the previous results on small-world networks.

In Fig. 3.4C we address the challenge in making causal statements about the relationship between network structure and dynamics – many network properties are inherently linked. As a result, the x-swap method also destroys other network

properties simultaneously. We compare four network properties in the original and x-swapped networks, and observe that the x-swap method affects clustering but also the ASPL, modularity, and assortativity. This means the changes in information flow seen in Fig. 3.4B may be due to changes in a combination of these (and possibly other) network properties.

3.2.3 Community structure and the weakness of long ties

The effects of long-range links on information flow have been investigated for some time, from Granovetter's seminal "strength of weak ties" observation in simple contagion [43] to Centola's contrasting "weakness of long ties" observation in complex contagion [42]. We use the stochastic block model described in Sec. 3.1.4 to elucidate this effect in the quoter model.

In Fig. 3.5 we measure information flow between and within groups. In Fig. 3.5A we see that as the within-block connection probability p_0 increases, the network density increases too, and thus it is not surprising that $h_\times$ increases. However, we note that the cross-entropy between blocks is always higher, supporting the weakness of long ties result. In Fig. 3.5B the difference in information flow $\Delta h_\times$ increases due to between-block links containing less and less predictive information. In Fig. 3.5C we see that this result is completely explained by the modularity, and not the individual connection probabilities p_0, p_1. A surprising result is that even when $p_0 < p_1$, the quantity $\Delta h_\times$ is positive. This suggests that even when there are no well-defined communities (i.e. a bipartite graph with a few extra links added in one bipartition)

within-block edges always contain more predictive information.

3.2.4 The role of dynamic heterogeneity

Previous research [75, 76] has highlighted the importance of including heterogeneity in the dynamical parameters of a model. We investigated how information flows in the stochastic block model when the nodes in the two blocks have different Zipf exponents: nodes in block A have exponent α_A and nodes in block B have exponent α_B.

Figure 3.6 shows how information flow changes when the two blocks have different vocabulary distributions (Fig. 3.6A,C) compared with the same distribution (Fig. 3.6B). For illustration, we show the Zipfian vocabulary distributions for the two groups as insets in Fig. 3.6. We observe a much larger trend in how cross-entropy changes with modularity when the exponents are not equal compared to when they are equal. This underscores how structural features (the degree of modularity) greatly magnifies the effects of intrinsic dynamic heterogeneity (different vocabulary distributions). While modularity plays a role even when the two groups have identical vocabulary distributions (Fig. 3.5), this difference is challenging to detect in Fig. 3.6B when viewed on the scale of groups with different vocabulary distributions (Fig. 3.6A,C).

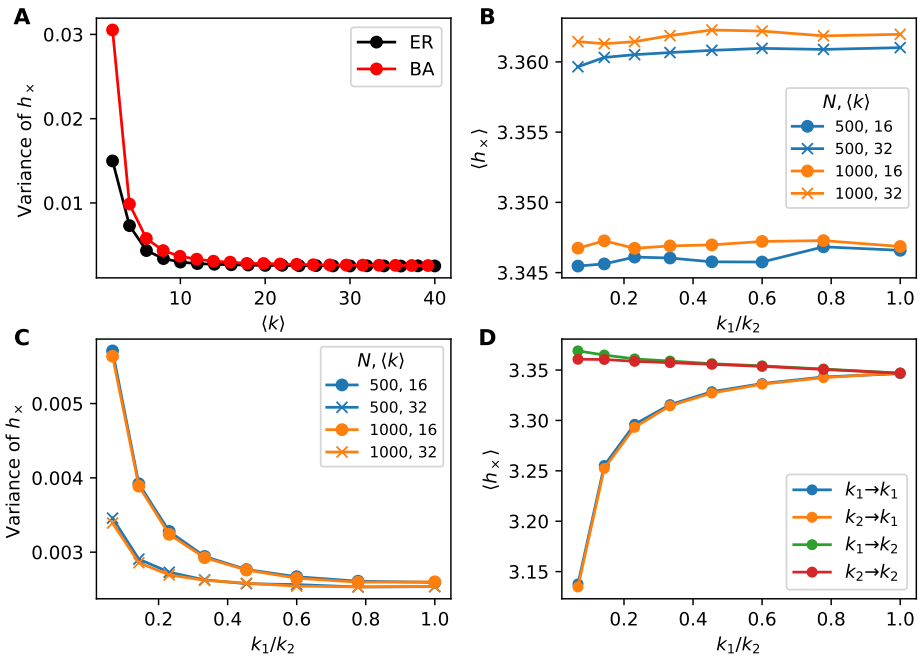

Figure 3.3: Exploring the variance of information flow. ***(A)*** *Variance of cross-entropy is higher at low densities for BA than ER networks despite the average $h_\times$ being similar (Fig. 3.1C).* ***(B-D)*** *Information flow on dichotomous networks (random networks where all nodes have degree k_1 or degree k_2, allowing tunable degree heterogeneity) of size $N \in \{500, 1000\}$ with $\langle k \rangle \in \{16, 32\}$. Each point constitutes 500 trials.* ***(B)*** *Average cross-entropy versus k_1/k_2. Degree heterogeneity does not affect average cross-entropy, supporting Fig. 3.1C. Network size has a smaller affect on $h_\times$ compared to the average degree.* ***(C)*** *Variance of cross-entropy versus k_1/k_2. Higher degree heterogeneity (lower k_1/k_2) leads to higher variation in $h_\times$ over links, indicating the existence of highly predictive nodes and nodes that contribute little predictive information within heterogeneous networks.* ***(D)*** *Dichotomous networks of size $N = 1000$ and $\langle k \rangle = 16$. Average cross-entropy over links conditioned on degrees of endpoints (predicting ego from alter). Only the degree of the ego matters, approximately, not the degree of the alter.*

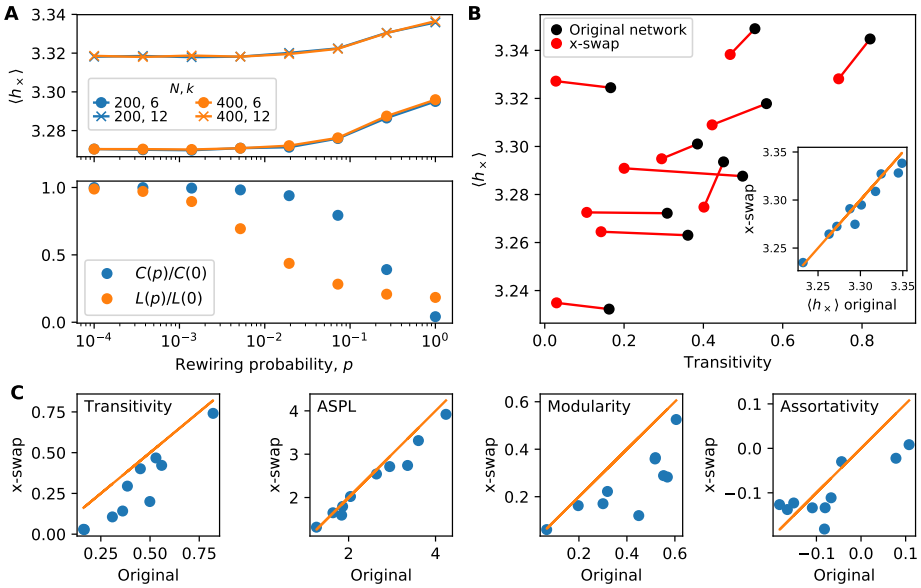

Figure 3.4: Mixed effects of clustering on information flow. ***(A)*** *Information flow on small-world networks of size $N \in \{200, 400\}$ and average degree $k \in \{6, 12\}$. As network rewiring increases (and clustering decreases) $h_\times$ increases. This suggests that clustered networks promote information flow. Rewiring a small-world network changes the diameter (L) as well the clustering (panel A, bottom); however, $h_\times$ begins to increase primarily when the clustering begins to drop, not when diameter begins to drop. Each point constitutes 300 trials.* ***(B)*** *Average cross-entropy versus transitivity for real-world networks. By randomizing networks using the standard "x-swap" method (Sec. 3.1.4), we can lower the transitivity and investigate how $h_\times$ changes. Some networks show little change in $h_\times$ on randomized networks compared with the original networks, while others show a slight decrease in $h_\times$. This is especially visible in the inset comparing $h_\times$ directly. Each point constitutes 300 simulations.* ***(C)*** *Several network properties before and after the x-swap method. While the x-swap method lowers transitivity, it also alters other important network properties, making it challenging to isolate the role of clustering from other properties.*

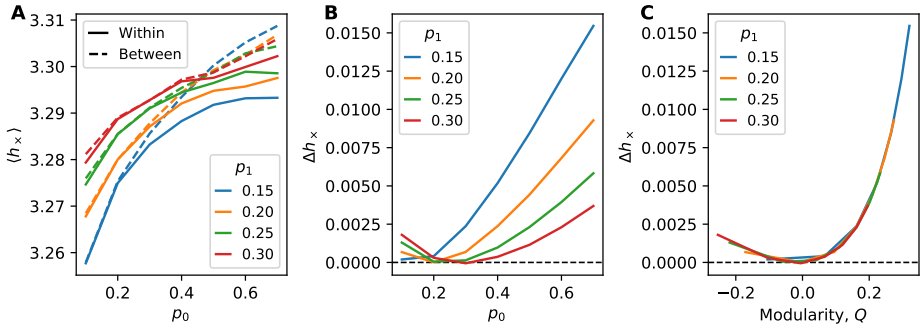

Figure 3.5: Information flow within the stochastic block model (SBM) of $N = 100$ (two blocks of size $N = 50$). Each point constitutes 10k trials. ***(A)*** *Average cross-entropy on within-block edges and between-block edges as a function of the within-block connection probability p_0 for different between-block connection probabilities p_1.* ***(B, C)*** *Examining the cross-entropy difference $\Delta h_\times \equiv \langle h_\times(between)\rangle - \langle h_\times(within)\rangle$ across* ***(B)*** *connection probabilities and* ***(C)*** *modularity Q. Examining $\Delta h_\times$ as a function of modularity Q shows a clear collapse across values of SBM probabilities. Interestingly, anti-community structure ($Q < 0$) still leads to positive $\Delta h_\times$, indicating that information flow is still more prevalent within blocks.*

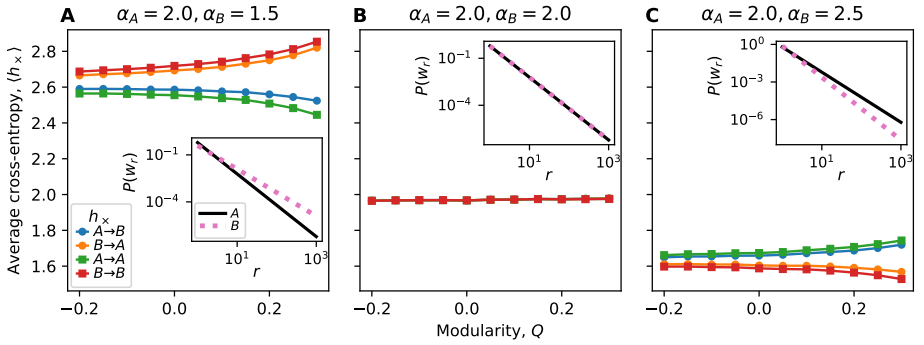

Figure 3.6: Effects of dynamic heterogeneity on information flow in the stochastic block model. Nodes in group A have Zipfian vocabulary distribution with exponent α_A while nodes in B have exponent α_B. The between-block connection probability is fixed ($p_1 = 0.15$) and the within-block connection probability p_0 is varied to generate a range of modularities. Since the structure is symmetric (subgraphs A and B have the same size and expected density), we only show the result of fixing $\alpha_A = 2$ and varying α_B. Each point constitutes 150 trials. ***(A)*** *The vocabulary distribution of group A has a lower Shannon entropy than of B, and this difference is visible from examining links $A \to A$ and $B \to B$. When examining links $A \to B$ and $B \to A$, the cross-entropy is mainly dependent on the vocabulary distribution of the alter. As modularity increases, differences between the predictabilities of various nodes are exaggerated.* ***(B)*** *In homogeneous communities, the cross-entropy does not vary with modularity at such a scale.* ***(C)*** *The vocabulary distribution of group A has a higher Shannon entropy than of B. Similar mirror results are seen as in panel A.*

Chapter 4

Analysis of the quoter model

In this chapter we present some analytical results for the quoter model. The quoter model in its general form is challenging to analyze for several reasons. First, it is nontrivial to write down as a stochastic process since we must keep track of the entire history of the process and the length of the process is variable (due to the message-length distribution). What exactly is a state, and how do we specify transition rates? Second, on arbitrary networks, we need to account for feedback loops and indirect influences (i.e. u quotes v and v quotes w, so information flows indirectly from u to w). Third, the quantity $h_\times$ (from the cross-entropy estimator) is difficult to compute unless the exact text is known. Even if we assume the length of the text generated at time t is exactly $t\lambda$, when computing Λ_t we need to consider all possible partitions of $(T-t)\lambda$ future words: (i) into messages of lengths $m_1, m_2, ...$ such that $\sum m_i = (T-t)\lambda$ and (ii) into types, quote or random, for each message. Nonetheless, if we make some simplifying assumptions and restrict ourselves to special cases of the quoter model, we will find interesting analytical results.

4.1 One link

Here we study the most basic network: two nodes u, v connected by one directed link (u, v). That is, v quotes u but u does not quote v. Approximate calculations of $h_\times$ for this scenario were given in [1]. We restate them here and offer some improvements. The probability that $\Lambda_t = \ell$ is obtained from the law of total probability.

$$\begin{aligned}\mathbb{P}(\Lambda_t = \ell) &= \mathbb{P}(\Lambda_t = \ell \mid \text{quote})\mathbb{P}(\text{quote}) + \mathbb{P}(\Lambda_t = \ell \mid \text{random})\mathbb{P}(\text{random}) \\ &\equiv q\mathbb{P}(\Lambda_Q = \ell) + (1-q)\mathbb{P}(\Lambda_R = \ell).\end{aligned}$$

In the first line, we condition on whether the message at position t was quoted or randomly generated. We then denote $\mathbb{P}(\Lambda_Q = \ell) = \mathbb{P}(\Lambda_t = \ell \mid \text{quote})$, where we have suppressed the dependence on the position t. Taking the expectation of both sides over ℓ, we get Λ_t, the expected longest match length at position t.

$$\Lambda_t = q\Lambda_Q + (1-q)\Lambda_R$$

with Λ_Q and Λ_R denoting the appropriate conditional expectations. We thus need to estimate Λ_Q and Λ_R.

Λ_R: Suppose t words have been posted by the ego in total (thus approximately t words have been posted by the alter also), and assume that the next message posted by the ego is randomly generated. We now ask, what is the probability that the next m words posted by the ego match somewhere in the alter's text? The probability

that the ego's next word matches a specific location in the alter's text is

$$d \equiv \sum_{i=1}^{z} W(w_i)^2$$

which is the Simpson index of the vocabulary distribution. There are $t - m + 1 \approx t$ (when $t \gg m$) positions in the alter's text where a match of length at least m can occur, and each occurs with probability d^m. Hence the expected number of matches of length at least m is td^m, which we denote $C(m)$. The expected longest match length m^* should satisfy $C(m^*) \geq 1$ and $C(m^* + 1) < 1$. That is, we will be likely to see a match of length m^* but not of length $m^* + 1$. Solving for $C(m^*) = 1$ we find

$$m^* = \frac{\log(t)}{\log(1/d)}.$$

Finally, Λ_R is one more than the longest match length.

$$\Lambda_R = \frac{\log(t)}{\log(1/d)} + 1.$$

Λ_Q: Again suppose the ego and alter have posted t words, and assume that the next message posted by the ego is a quote of length λ. Then, clearly $\Lambda_t \geq \lambda + 1$. What is the expected amount that random chance extends this match length? There are $(t - \lambda)d^\lambda + 1 \approx td^\lambda + 1$ copies of the original quote in the alter's text including the original quote itself. The expected number of matches of length at least m which follow any copy of the quote, is then $(td^\lambda + 1)d^m \equiv C(m)$. Solving $C(m^*) = 1$ as before, we get

$$m^* = \frac{\log(td^\lambda + 1)}{\log(1/d)}.$$

It is not accurate to say that Λ_Q equals $\lambda + m^* + 1$. If $\lambda + m^* + 1 \gg \Lambda_R$ then it is likely that the match is solely due to the quote at position t. We then expect $\Lambda_{t+1} = \Lambda_t - 1$, $\Lambda_{t+2} = \Lambda_t - 2$, and so on until random matching is again likely. Based on an observation from [1], we find that Λ_Q decays linearly back down to Λ_R. We thus derive the average Λ_Q at position t to be

$$\Lambda_Q = \frac{1}{\tilde{\lambda} - \Lambda_R + 2} \sum_{j=0}^{\tilde{\lambda}-\Lambda_R+1} (\tilde{\lambda} - j + 1) = \frac{1}{2}\left[\lambda + \frac{\log(td^\lambda + 1)}{\log(1/d)} + \frac{\log(t)}{\log(1/d)} + 2\right]$$

with $\tilde{\lambda} = \lambda + \frac{\log(td^\lambda+1)}{\log(1/d)}$.

We now use the formula $\Lambda_t = (1-q)\Lambda_R + q\Lambda_Q$ and approximate $\Lambda \equiv \sum_t \Lambda_t$ by an integral.

$$\begin{aligned}
\Lambda &= \sum_{t=1}^{T} \Lambda_t \\
&\approx \int_0^T \{(1-q)\Lambda_R + q\Lambda_Q\}\, dt \\
&= \int_0^T \left\{(1-q)\left[\frac{\log(t)}{\log(1/d)} + 1\right] + q \cdot \frac{1}{2}\left[\lambda + \frac{\log(td^\lambda + 1)}{\log(1/d)} + \frac{\log(t)}{\log(1/d)} + 2\right]\right\} dt \\
&= \frac{T}{\log(1/d)}\left\{(1-q)\left(\log\frac{T}{d} - 1\right) + \frac{q}{2}\left[\log\frac{T}{d^{\lambda+2}} + \left(\frac{1}{Td^\lambda} + 1\right)\log(Td^\lambda + 1) - 2\right]\right\}
\end{aligned}$$

The cross-entropy $h_\times$ can be computed by substituting Λ into the estimator in Sec. 2.4. By a standard asymptotic analysis we find

$$\lim_{T\to\infty} h_\times = \log(1/d).$$

This expression is equivalent to the Rényi entropy of the vocabulary distribution

$$h_\alpha = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{z} W(w_i)^\alpha \right)$$

with $\alpha = 2$. The Rényi entropy with $\alpha = 2$ is always lower than the Shannon entropy, which is the limiting case as $\alpha \to 1$. One might think that $h_\times$ would converge to the Shannon entropy of the vocabulary distribution, since the frequency of words in the alter's and ego's texts will converge to the common vocabulary distribution, and the cross-entropy $H(W, W)$ is equivalent to the Shannon entropy. This indicates that there is reduced uncertainty due to the time ordering, and thus the quoter model has interesting temporal features which the estimator used here captures because it satisfies temporal precedence. The Rényi entropy also appears in the longest common subsequence problem with non-iid sequences [77]. Shockingly, the limiting behavior of the cross-entropy only depends on d and does not depend on q or λ. However, for finite time we shall find that $h_\times$ does exhibit noticeable dependence on q and λ. Figure 4.1 shows qualitative agreement between theoretical and actual values of $h_\times$. There are some discrepancies at extreme values of q, which appear to be exaggerated at higher values of z (higher diversity vocabularies). We see that $h_\times$ decreases as q and λ increase. On the other hand $h_\times$ increases with z. This is intuitive based on the definition of the estimator; more frequent, longer copying, in concert with a lower diversity vocabulary leads to long match lengths and ultimately lower $h_\times$.

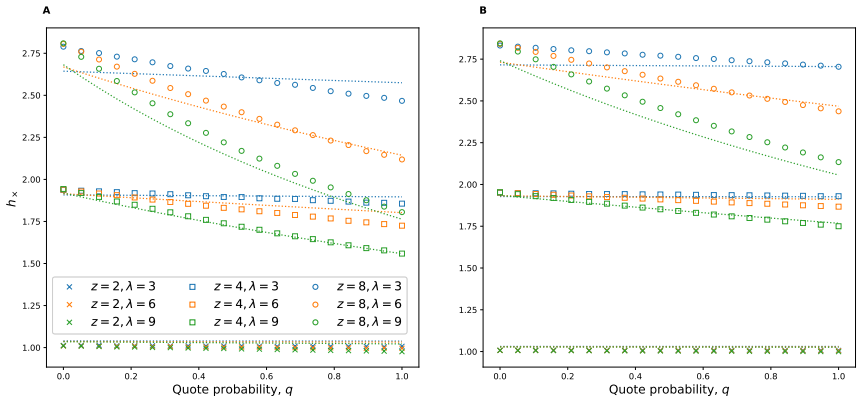

*Figure 4.1: Theoretical (lines) and empirical (symbols) values of $h_\times$ in the quoter model for varying parameter values q, λ, z. This figure has been adopted from [1]. Symbols represent empirical $h_\times$ averaged over 100 simulations. The vocabulary distribution is discrete uniform on $\{1, \ldots, z\}$. **(A)** $T = 1000$, so that the length of the text is $\approx 1000\lambda$, **(B)** $T = 10000$, so that the length of the text is $\approx 10000\lambda$*

4.2 New contributions

Here we provide fresh insights into the approximations presented in the previous section. Firstly, in calculating Λ_R in Sec. 4.1 it was implicitly assumed, when solving for m^*, that the remainder (or at least the immediate future) of the ego's text was entirely random. Suppose that there are $t = 1500$ alter words prior to the ego's current word and $d = 1/2$, then $m^* = \frac{\log 1500}{\log 2} \approx 10.55$ which extends 3+ messages of length $\lambda = 3$. Unless q is very low, it is likely that the ego makes a quote within these next 3 messages. If quoting at the second message, there is a slightly higher probability of extending the previous match of length m, than if the message was random. This is because the quote can come immediately after (any copy of) the previous match in the alter's text. This occurs roughly with probability d^m since there are $\approx td^m$ copies of this match and the probability of quoting at any particular

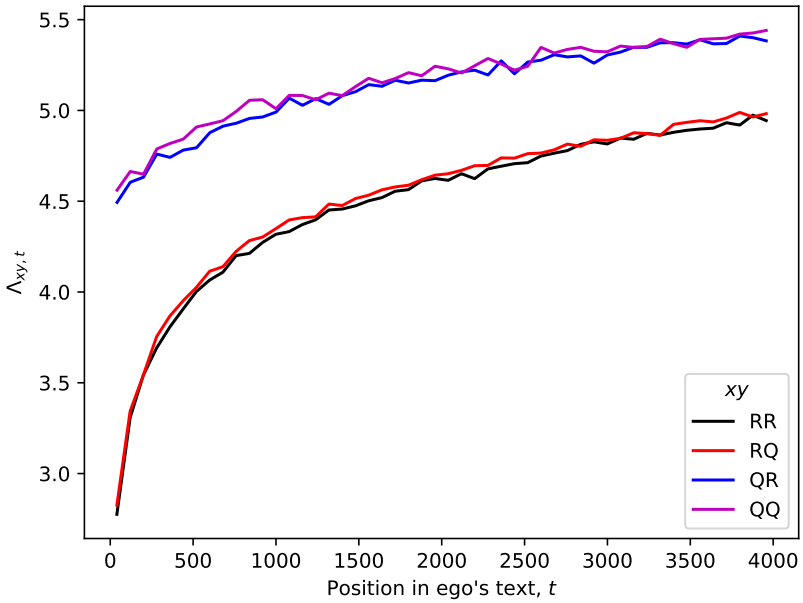

Figure 4.2: The match length Λ_t given that the next two messages are (R)andomly generated or (Q)uoted. When conditioning on the status of next two messages, we call this a "2nd order approximation of Λ_t." The parameters for the simulations are $q = 1/2$, $\lambda = 3$, and $T = 2000$. The vocabulary distribution is discrete uniform on $\{1, \ldots, z\}$ with $z = 8$. There are 300 simulations and the curves are the result of binning t and computing an average Λ_t in each bin

position is $1/t$. In calculating Λ_Q we similarly assumed the immediate future after the quote was randomly generated. Although this seems like a naive approximation, Fig. 4.2 suggests that the benefit from further partitioning the ego's future text into R and Q is negligible. We see that Λ_t primarily only depends on the whether the first message is random or quoted. Also, when the second message is a quote, Λ_t is only slightly larger as we expected. This (non-) result motivates us to look elsewhere for improvements.

It turns out, solving $C(m^*) = 1$ in calculating Λ_R and Λ_Q is only a heuristic, for which there exist better approximations. Two well-studied problems in probability

theory which bear a resemblance to ours are the "longest common substring" problem [77, 78] and the "longest run of heads" problem [79, 80]. In the former, given two random sequences of length n we want to calculate the expected longest consecutive subsequence which appears in both sequences. This problem has direct applications to DNA sequence comparison – with the longest common substring being a measure of similarity between two DNA sequences. A simpler problem is: suppose a fair coin is flipped n times, what is the expected length of the longest run of heads, R_n?. Note that a run of heads is necessarily preceded by a tail, which occurs with probability $1/2$. Thus the expected number of "head runs" is $n/2$. We can model each independent head run as a geometric random variable with parameter $p = 1/2$. The longest head run R_n is then given by the maximum of $n/2$ iid geometric random variables. Although there here is no explicit formula for the maximum of geometric random variables [81], an approximation leads to [79, 80]

$$R_n \approx \log_2(n/2) + \frac{\gamma}{\log 2} - \frac{1}{2}.$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. This can easily be generalized to the case of a biased coin with probability of heads p and tails $q = 1 - p$. We have

$$R_n \approx \frac{\log(nq)}{\log(1/p)} + \frac{\gamma}{\log(1/p)} - \frac{1}{2}.$$

We now make the connection to the quoter model. The longest match length starting at position t is equivalent to the longest run of heads in t tosses (each toss corresponding to a position in the alter's text), where the probability of a head is $d = \sum_{i=1}^{z} W(w_i)^2$. However, a match does not need to be preceded by a non-match.

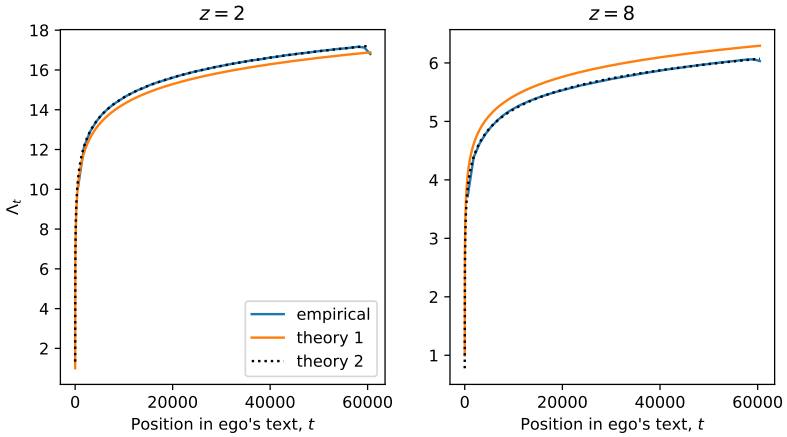

Figure 4.3: Old (theory 1) and new (theory 2) approximations for Λ_R at position t. There is very good agreement between the new approximation and empirical values. The parameters for the quoter model are $q = 0$, $\lambda = 3$, and $T = 20000$. The vocabulary distribution is discrete uniform on $\{1, \ldots, z\}$ with $z \in \{2, 8\}$. There are 500 simulations and the curves are the result of binning t and computing an average Λ_t in each bin

That is, every position t in the alter's text is capable of being the start of a match, regardless of their previous text. This contrasts with the coin-flipping example, in which if a heads is seen previously then the current position cannot be the start of a run and if a tails is seen previously then the current position must mark the start of a run. This means the term $\log(nq)/\log(1/p)$ becomes $\log(n)/\log(1/d)$. We now have

$$\Lambda_R = \frac{\log t}{\log(1/d)} + \frac{\gamma}{\log(1/d)} + \frac{1}{2}$$

where we have added 1 because Λ is one more than the longest match. This is precisely the expression we had before, but with the added constants $\frac{\gamma}{\log(1/d)} - \frac{1}{2}$. In Fig. 4.3 we show the improvement in calculating Λ_R when $q = 0$.

Similar reasoning suggests that the approximation for Λ_Q can be improved by including these constants. That is, in calculating Λ_Q we had $td^\lambda + 1$ iid geometric

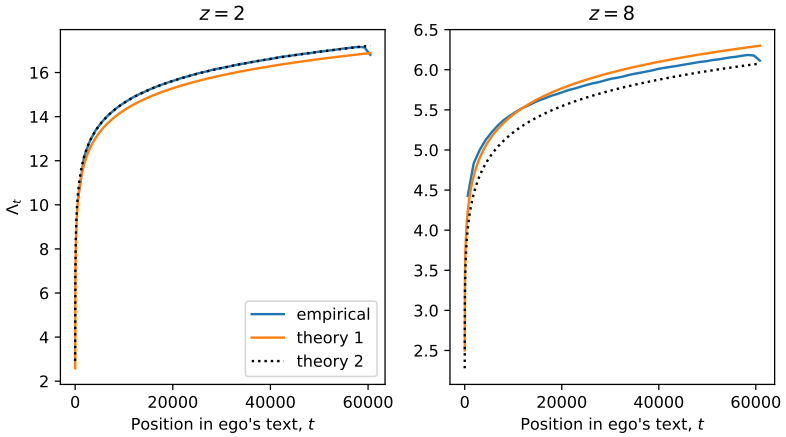

Figure 4.4: Old (theory 1) and new (theory 2) approximations for Λ_Q at position t. There is improved agreement between the approximation and empirical values when $z = 2$, but notable error when $z = 8$ (when the vocabulary is more diverse). The parameters for the quoter model are $q = 1$, $\lambda = 3$, and $T = 20000$. The vocabulary distribution is discrete uniform on $\{1, \ldots, z\}$ with $z \in \{2, 8\}$. There are 500 simulations and the curves are the result of binning t and computing an average Λ_t in each bin

random variables. Thus Λ_Q should be

$$\Lambda_Q = \frac{1}{2}\left[\lambda + \frac{\log(td^\lambda + 1)}{\log(1/d)} + \frac{\log(t)}{\log(1/d)} + 2\right] + \frac{\gamma}{\log(1/d)} - \frac{1}{2}.$$

In Fig. 4.4 we see that this is an improved approximation when $z = 2$, but there is noticeable error when $z = 8$. Both approximations appear to be growing faster than the empirical Λ_Q, implying there is error which is not due to the additional constants. Therefore it is hard to judge whether or not these constants should be included. We subsequently will choose to use the old approximation for Λ_Q in calculating $h_\times$.

Finally, in Fig. 4.5 we show the impact of these new approximations on the overall quantity of interest, $h_\times$. There is nearly exact agreement when q is small, as we expected based on Fig. 4.3. However, when q is large there are discrepancies which seem to be exaggerated as higher values of z, coinciding with our observations in

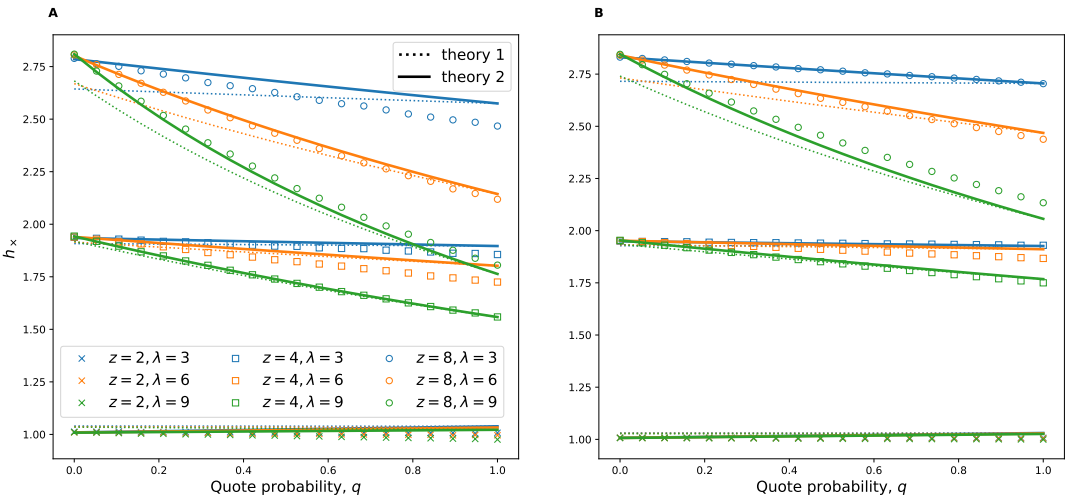

*Figure 4.5: Comparing the old (theory 1) and new (theory 2) approximations for $h_\times$ for varying parameter values q, λ, z. The newly developed approximation shows slight improvement for some parameter combinations. Symbols represent empirical $h_\times$ averaged over 100 simulations. The vocabulary distribution is discrete uniform on $\{1, \ldots, z\}$. **(A)** $T = 1000$, so that the length of the text is $\approx 1000\lambda$, **(B)** $T = 10000$, so that the length of the text is $\approx 10000\lambda$.*

Fig. 4.4.

We briefly address some possible future improvements. Including a "second order" approximation as discussed earlier, may help to improve this approximation. Using the law of total probability, one would arrive at an expression

$$\begin{aligned}\mathbb{P}(\Lambda_t = \ell) = \sum_{m_1}\sum_{m_2} \frac{\lambda^{m_1+m_2}e^{-2\lambda}}{m_1!m_2!}\Big(&\mathbb{P}(\ell|Q,Q,m_1,m_2)q^2 \\ &+ \mathbb{P}(\ell|R,R,m_1,m_2)(1-q)^2 + \mathbb{P}(\ell|Q,R,m_1,m_2)q(1-q) \\ &+ \mathbb{P}(\ell|R,Q,m_1,m_2)(1-q)q\Big)\end{aligned}$$

where for instance $\mathbb{P}(\ell|R,Q,m_1,m_2)$ is the probability $\Lambda_t = \ell$ given that the next two messages are random and then quoted, and of lengths m_1, m_2 respectively. This

could be simplified by assuming $m_1 = m_2 = \lambda$ and then one only needs to calculate $\mathbb{P}(\ell|R, Q)$. However, it is still not obvious how to calculate these terms. If the next two messages are both of length λ, one must consider the probability that the match ends before the first message, between the first and second messages, or after the second message. Thus $\mathbb{P}(\ell|R, Q)$ would be defined piecewise and require calculating the probability of each of these three cases. From the condition $\frac{\log t}{\log(1/d)} < \lambda$ we can determine $t < \frac{e^\lambda}{d}$. That is, for t sufficiently small it is unlikely that the longest match extends more than the first message. Another place for improvement would be to make rigorous our heuristic argument for incorporating the observed decay of Λ_Q following the start of a quote. Lastly, we neglected the fact that the alter's text is growing simultaneously, and so a match could arise within the extended text; however, this should only have a minor effect. In summary, the quoter model in a simple setting (one directed link) provides an interesting and challenging theoretical problem which has ties to other well-studied problems in probability theory. Although our approximations do not show significant improvement, we have at least put this problem on a more rigorous footing and identified places for improvement.

4.3 Connection to the voter model

One may criticize the quoter model for being too novel, both in its dynamical rules and observable, $h_\times$. We now argue that this is not the case, and in fact the dynamical rules of the quoter model are roughly a generalization of a voter model variant.

Suppose that the message length distribution is $p(\lambda') = \delta(\lambda' - \lambda)$, where $\lambda > 0$ is fixed. Furthermore let quotes be taken from the immediate past rather than the

entire past. That is, when copying, individuals will always quote the last λ words of their neighbor. Notice that the λ words in each message are all independent, i.e. the first word is only ever copied from the first word of a neighbor, and not influenced by words two, three, etc. That is, the nodes are essentially voting on what the first word of the message should be, and separately voting on what the second word of the message should be, and so on for all λ words. Note this special case highlights the simplicity of our language model – there is no correlation between words. We can thus represent this case as λ iid realizations of a noisy voter model with multiple states [82]. Here there are z states corresponding to the z unique words, and the level of noise is determined by $1 - q$ plus the vocabulary distribution.

We can learn a lot from looking at one realization since, by independence, statistics of the λ realizations can be calculated as a product. We momentarily drop our study of the cross-entropy and focus on the distribution of words (i.e. the fraction of individuals using word i at time t). Much research has been done on the noisy voter model and its extensions. The model is very general and has attracted researchers from finance, biology, and physics. The remainder of this section will be devoted to reviewing the noisy voter model. We specifically adapt the noisy voter model for our context and provide complete derivations.

The noisy voter model is a Markov process, usually described in continuous time. The process is specified by the initial number of nodes in each state, and by transition rates. Let $x_i;\ i = 1, \ldots, z$ be the fraction of nodes using word i, which depends on time and is the typical quantity of interest in voter models. Let $\epsilon_i \equiv W(w_i)$ be the probability of word i via the vocabulary distribution. Consider the case of the complete graph. Let $\delta \equiv 1/N$. In a single update event we have either $x_i \to x_i + \delta$

(x_i raises by δ) or $x_i \to x_i - \delta$ (x_i lowers by δ). The transition $x_i \to x_i + \delta$ occurs with rate

$$\begin{aligned} R_i(\mathbf{x}) &= N \sum_{j \neq i} (q x_j x_i + (1-q)\epsilon_i x_j) \\ &= N[q(1-x_i)x_i + (1-q)\epsilon_i(1-x_i)] \end{aligned}$$

where $\mathbf{x} \equiv (x_1, \ldots, x_z)$. The first term in the sum is the contribution when quoting: $x_i x_j$ is the probability of first choosing a node using word $j \neq i$ and a neighbor using word i to copy. The second term is the contribution when not quoting: x_j is the probability of choosing a node using a word $j \neq i$, then the probability of using word i according to their vocabulary distribution is ϵ_i. The factor of N in front is a conventional time rescaling that ensures $O(N)$ events per unit time. Similarly the transition $x_i \to x_i - \delta$ occurs with rate

$$\begin{aligned} L_i(\mathbf{x}) &= N \sum_{j \neq i} (q x_i x_j + (1-q)\epsilon_j x_i) \\ &= N[q(1-x_i)x_i + (1-q)(1-\epsilon_i)x_i] \end{aligned}$$

Notice that the rates only depend on i and therefore we can define a closed birth-death process for each species (word).

The master equation [83, 84] describing the evolution of $\mathbb{P}(x_i, t)$ is

$$\begin{aligned} \frac{\partial \mathbb{P}(x_i,t)}{\partial t} &= R_i(x_i - \delta)\mathbb{P}(x_i - \delta, t) + L_i(x_i + \delta)\mathbb{P}(x_i + \delta, t) \\ &\quad - [R_i(x_i) + L_i(x_i)]\mathbb{P}(x_i, t). \end{aligned}$$

The first two terms describe the inflow to state x_i, where the last term is the outflow

to other states.

Following a similar calculation as in [85] we expand to the second order in δ to obtain the Fokker–Planck equation (forward Kolmogorov equation). The Fokker-Planck equation is the continuous analog to the master equation, in which the probability mass function $\mathbb{P}(x_i, t)$ becomes a probability density function. To help with readability, we suppress the index i and the dependence of R, L, $\mathbb{P}$ on x_i, t. All derivatives on the right-hand side are partials with respect to x_i.

$$\begin{aligned}
\frac{\partial \mathbb{P}}{\partial t} &= \left(R - \delta R + \frac{\delta^2}{2} R''\right)\left(\mathbb{P} - \delta \mathbb{P}' + \frac{\delta^2}{2}\mathbb{P}''\right) \\
&\quad + \left(L + \delta L + \frac{\delta^2}{2} L''\right)\left(\mathbb{P} + \delta \mathbb{P}' + \frac{\delta^2}{2}\mathbb{P}''\right) \\
&\quad - (R+L)\mathbb{P} \\
&= -\delta\left[(R-L)\mathbb{P}' + (R-L)'\mathbb{P}\right] + \frac{\delta^2}{2}\left[(R+L)\mathbb{P}'' + 2(R+L)'\mathbb{P}' + (R+L)''\mathbb{P}\right] \\
&= -\frac{\partial}{\partial x}\left[\delta(R-L)\mathbb{P}\right] + \frac{\partial^2}{\partial x^2}\left[\frac{\delta^2}{2}(R+L)\mathbb{P}\right] \\
&\equiv -\frac{\partial}{\partial x}(v\mathbb{P}) + \frac{\partial^2}{\partial x^2}(D\mathbb{P})
\end{aligned}$$

where

$$\begin{aligned}
v(x_i) &= \delta x_i \left[R_i(x_i) - L_i(x_i)\right] \\
&= (1-q)(\epsilon_i - x_i) \\
D(x_i) &= \frac{\delta x_i^2}{2}\left[R_i(x_i) + L_i(x_i)\right] \\
&= \frac{1}{2N}\left[2q(1-x_i)x_i + (1-q)(\epsilon_i + (1-2\epsilon_i)x_i)\right]
\end{aligned}$$

are the drift and diffusion coefficients. We could have also derived this equation immediately from a Kramers-Moyal expansion of the master equation [83,84]. The Fokker-Planck equation cannot be solved analytically. However, we can derive the stationary state probability distribution $\mathbb{P}(x_i)$. In the stationary state, one has $\partial\mathbb{P}/\partial t = 0$. This implies

$$\begin{aligned}
0 &= -\frac{\partial}{\partial x}(v\mathbb{P}) + \frac{\partial^2}{\partial x}(D\mathbb{P}) \\
0 &= (-v'\mathbb{P} - v\mathbb{P}') + (D''\mathbb{P} + 2D'\mathbb{P}' + D\mathbb{P}'') \\
0 &= \mathbb{P}'' + \frac{-v + 2D'}{D}\mathbb{P}' + \frac{-v' + D''}{D}\mathbb{P}.
\end{aligned}$$

The above is a second-order linear homogeneous ODE with varying coefficients. The solution is given by [82,86]

$$\mathbb{P}(x_i) = C_i \exp\left(-\int_0^{x_i} \frac{-v(y) + D'(y)}{D(y)} dy\right)$$

where C_i is a normalization constant. Using numerical integration we plot the stationary distribution for various values of q, N, and α, with $\epsilon_i \sim \text{Zipf}(1000, \alpha)$ as in the quoter model simulations. Since there are $i = 1, \ldots, 1000$ marginals, we only show the marginals corresponding to the rank 1 and rank 2 words. Words of lower rank have distributions similar to the rank 2 word. Figure 4.6 shows the well-known transition of the stationary distribution $\mathbb{P}(x_i)$ from a unimodal to bimodal shape. For low values of N or high values of q we see a bimodal shape in word 1, roughly meaning that either nobody or everyone is using word 1. For high values of N or low values of q we see a unimodal shape, where the most probable state involves a nonzero fraction of people using word 1. For word 2, the most probable state is that nobody

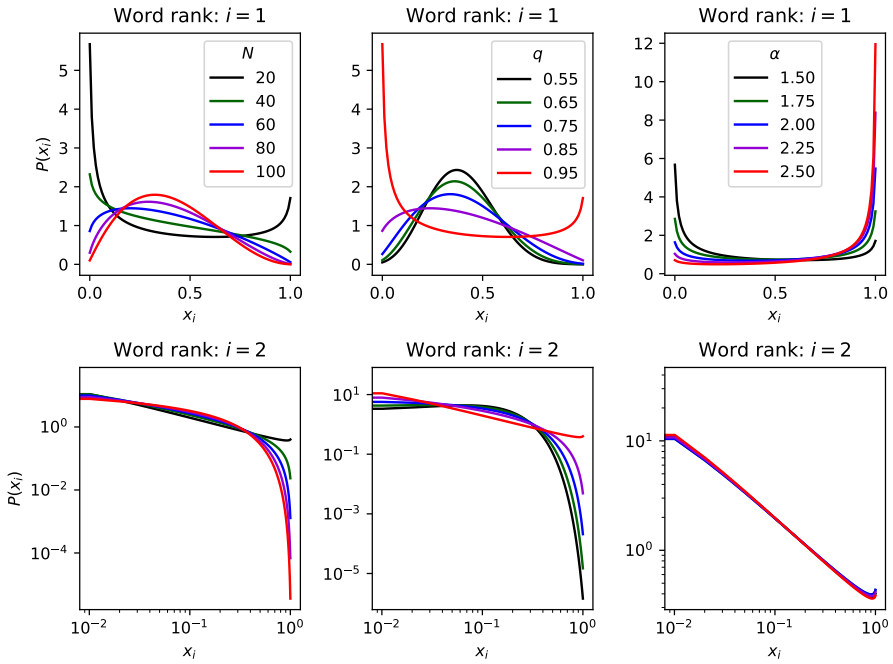

Figure 4.6: Marginal stationary distributions $\mathbb{P}(x_i)$ of the noisy voter model on the complete graph K_N. The distribution $\mathbb{P}(x_i)$ gives the probability that the fraction of nodes using word i is x_i in the stationary state. Parameter ranges are chosen to illustrate the transition of $\mathbb{P}(x_i)$ from a unimodal to bimodal shape. In the left column, N varies and $q = 0.95$, $\alpha = 1.5$. In the middle column, q varies and $N = 20$, $\alpha = 1.5$. In the right column, α varies and $N = 20$, $q = 0.95$. The second row is plotted on a log-log scale.

is using it. We conjecture that the approximate power-law regimes are the result of the power-law noise (from the vocabulary distribution), which to our knowledge has not been implemented in other studies of the noisy voter model.

Chapter 5

Discussion

In this thesis we studied a recent, information-theoretic approach to measuring and modeling information flow. Previous work has not fully utilized the massive amount of data available. Similar work has looked at the timings of tweets, but has not looked at the text itself. The cross-entropy estimator (Eq. 2.2) gives a measure of information flow which can incorporate the full text data available through online social media platforms, while simultaneously satisfying temporal precedence. We use the quoter model to test hypotheses about how information flows in social networks and to what extent the network structure mediates the flow of information.

In Chapter 3 we studied the quoter model on complex networks. We used random graph models and real-world networks to test how the degree distribution, clustering, and modularity play a role in information flow. We found that increased density inhibits information flow, which is also observed in traditional complex contagion models. In studying the importance of clustering, we simulated the quoter model on Watts-Strogatz networks and found that clustering promoted information flow. However, when applying a network randomization algorithm to real-world networks,

we found clustering often decreased information flow. These mixed results on the role of clustering in spreading align with the mixed results from complex contagion approaches. Ultimately, it is challenging to determine the effect of clustering on spreading because it is inherently confounded by other network properties such as density, average shortest path length, and modularity. Another result we found was that the quoter model replicated the "weakness of long ties" result from complex contagion. That is, ties that bridge communities tend to decrease information flow. Finally, we briefly studied the importance of parametric heterogeneity. In summary, despite not defining a thresholding/reinforcement mechanism, the quoter model, in combination with the cross-entropy, can produce features similar to those of complex contagion. This study suggests a more nuanced, information-theoretic measure of information flow can complement the study of information flow just as well as the increased complexity of models.

We mention some limitations of the study presented in Chapter 3. We only considered undirected, unweighted networks. In the context of social networks, this implies all relationships are reciprocal and equal in strength. Future work should extend to directed, weighted networks. Also, we only studied information flow by examining pairs of text streams. Information could flow from u to v and v to w, but not directly from u to v. The cross-entropy $h_{\times}(w \mid u)$ would be lower than expected, but u does not directly influence w. To address this, we could apply an analog of causation entropy to identify spurious flows [56], although this introduces the issue of needing to estimate high dimensional conditional entropies [1]. Lastly, while we observed a number of features that are signatures of complex contagion, there were features that were not revealed by the quoter model. Specifically, in [87] it was shown that there

is an optimal modularity Q which maximizes spreading in the threshold model.

In Chapter 4 we first study the quoter model on a simple network – one directed link. We review the calculations for $h_\times$ presented in [1]. We find this simple case has connections to the "longest run of heads" problem studied in probability theory. Borrowing from this line of research we were able to improve the calculations for Λ_R, and ultimately $h_\times$, but still noted discrepancies in Λ_Q. Future research should look to improve Λ_Q, possibly by a "second order" approximation. Also, a feasible next step may be to extend the analysis to larger networks, such as a three node "fork" or "collider." Secondly, we noted a connection between simplified case of the quoter model and the noisy voter model. We adapted prior calculations to our scenario and observed the well-known transition of the stationary distribution from a unimodal to bimodal shape. It is unclear how, if at all, these parameter regimes exist in the quoter model when quotes can be taken from the entire history of a neighbors text. This connection warrants further research.

We briefly mention a few other areas where our information-theoretic approach can be improved. Information-theoretic measures, such as the cross-entropy are inherently model-free which means no assumptions are made about how two individuals are related. The result is, we know we can use Alice's past text to help predict Bob's future text, but we do not know *how* to predict Bob's future text. After addressing this shortcoming, it may be possible to apply our approach to predict online social media activity. This is of great intrinsic interest to researchers at Facebook, Twitter, etc., but should also prompt concerns of data privacy and algorithmic personalization. Additional work may also make the quoter model a more realistic language model, i.e., by introducing correlations between words, or may seek to fit model parameters

to real-world data. It may be valuable to translate previous results on information flow, such as influence maximization, to this new approach.

In conclusion, this modern measure and model exhibits many desirable properties of preexisting models for social contagion and information flow, while taking into account the massive amount of text data available. We have also explored and revealed several exciting new properties not known by researchers in the field. Future research should continue testing hypotheses about information flow in social networks and validate these findings with real-world data. We note that along with this new approach comes the need for new analytical techniques and more efficient estimators. Following this approach, there are still many interesting open questions (both theoretical and practical) for future research on information flow to address.

Bibliography

[1] J. P. Bagrow and L. Mitchell. The quoter model: A paradigmatic model of the social flow of written information. *Chaos*, 28(7):075304, 2018.

[2] S. Kemp. Digital 2020: 3.8 billion people use social media. Jan. 2020. https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media, Accessed: 2020-03-04.

[3] J. Coleman, E. Katz, and H. Menzel. The diffusion of an innovation among physicians. *Sociometry*, 20(4):253–270, 1957.

[4] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5–es, 2007.

[5] A. Susarla, J.-H. Oh, and Y. Tan. Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, 23(1):23–41, 2012.

[6] L. Gao, C. Song, Z. Gao, A.-L. Barabási, J. P. Bagrow, and D. Wang. Quantifying information flow during emergencies. *Scientific Reports*, 4:3997, 2014.

[7] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.

[8] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.

[9] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1):35, 2019.

[10] N. Perra and L. E. Rocha. Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific Reports*, 9(1):1–11, 2019.

[11] L. Hébert-Dufresne, S. V. Scarpino, and J.-G. Young. Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement. *Nature Physics*, pages 1–6, 2020.

[12] B. Min and M. San Miguel. Competing contagion processes: Complex contagion triggered by simple contagion. *Scientific Reports*, 8(1):1–8, 2018.

[13] J. M. Epstein, J. Parker, D. Cummings, and R. A. Hammond. Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PLoS One*, 3(12), 2008.

[14] S. Funk, M. Salathé, and V. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.

[15] J. Borge-Holthoefer, R. A. Baños, S. González-Bailón, and Y. Moreno. Cascading behaviour in complex socio-technical networks. *Journal of Complex Networks*, 1(1):3–24, 2013.

[16] D. Guilbeault, J. Becker, and D. Centola. Complex contagions: A decade in review. In *Complex spreading phenomena in social systems*, pages 3–25. Springer, 2018.

[17] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440, 1998.

[18] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008(10):P10008, 2008.

[20] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. *Journal of Statistical Mechanics*, 2005(09):P09008, 2005.

[21] B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[22] M. E. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.

[23] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

[24] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.

[25] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591, 2009.

[26] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang. A survey on the fusion process in opinion dynamics. *Information Fusion*, 43:57–65, 2018.

[27] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A.*, 115(772):700–721, 1927.

[28] H. Abbey. An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24(3):201, 1952.

[29] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[30] D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[31] D. Centola, V. M. Eguíluz, and M. W. Macy. Cascade dynamics of complex propagation. *Physica A*, 374(1):449–456, 2007.

[32] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[33] P. S. Dodds and D. J. Watts. Universal behavior in a generalized model of contagion. *Physical Review Letters*, 92(21):218701, 2004.

[34] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 67–75. Springer, 2008.

[35] M. E. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[36] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[37] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(4):1–37, 2012.

[38] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.

[39] D. J. O'Sullivan, G. J. O'Keeffe, P. G. Fennell, and J. P. Gleeson. Mathematical modeling of complex contagion on clustered networks. *Frontiers in Physics*, 3:71, 2015.

[40] J. C. Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.

[41] C. Gray, L. Mitchell, and M. Roughan. Super-blockers and the effect of network structure on information cascades. In *Companion Proceedings of the The Web Conference 2018*, pages 1435–1441, 2018.

[42] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734, 2007.

[43] M. S. Granovetter. The strength of weak ties. In *Social networks*, pages 347–367. Elsevier, 1977.

[44] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[45] M. O. Jackson. *Social and economic networks*. Princeton University Press, 2010.

[46] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[47] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, pages 643–663, 1975.

[48] S. Redner. Reality-inspired voter models: A mini-review. *Comptes Rendus Physique*, 20(4):275–292, 2019.

[49] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226, 1997.

[50] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[51] T. Weissman. EE 376A: Information theory lecture notes, Jan. 2016. https://web.stanford.edu/class/ee376a/, Accessed: 2020-02-16.

[52] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[53] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461, 2000.

[54] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, pages 424–438, 1969.

[55] L. Barnett, A. B. Barrett, and A. K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical Review letters*, 103(23):238701, 2009.

[56] J. Sun and E. M. Bollt. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D*, 267:49–57, 2014.

[57] J. Sun, D. Taylor, and E. M. Bollt. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1):73–106, 2015.

[58] M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Physical Review Letters*, 100(15):158101, 2008.

[59] C. E. Shannon. Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1):50–64, 1951.

[60] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov, and A. J. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Transactions on Information Theory*, 44(3):1319–1327, 1998.

[61] J. Ziv and N. Merhav. A measure of relative entropy between individual sequences with application to universal classification. *IEEE Transactions on Information Theory*, 39(4):1270–1279, 1993.

[62] J. P. Bagrow, X. Liu, and L. Mitchell. Information flow reveals prediction limits in online social activity. *Nature Human Behaviour*, 3(2):122–128, 2019.

[63] J. Borge-Holthoefer, N. Perra, B. Gonçalves, S. González-Bailón, A. Arenas, Y. Moreno, and A. Vespignani. The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Science Advances*, 2(4):e1501158, 2016.

[64] G. Ver Steeg and A. Galstyan. Information transfer in social media. In *Proceedings of the 21st International Conference on World Wide Web*, pages 509–518, 2012.

[65] G. Ver Steeg and A. Galstyan. Information-theoretic measures of influence based on content dynamics. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pages 3–12, 2013.

[66] A. De, I. Valera, N. Ganguly, S. Bhattacharya, and M. G. Rodriguez. Learning and forecasting opinion dynamics in social networks. In *Advances in Neural Information Processing Systems*, pages 397–405, 2016.

[67] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[68] T. Pond, S. Magsarjav, T. South, L. Mitchell, and J. P. Bagrow. Complex contagion features without social reinforcement in a model of social information flow. *Entropy*, 22(3):265, 2020.

[69] J. Miller and T. Ting. EoN (Epidemics on networks): A fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, 4(44):1731, 2019.

[70] R. Lambiotte. How does degree heterogeneity affect an order-disorder transition? *Europhysics Letters*, 78(6):68002, 2007.

[71] A. Carro, R. Toral, and M. San Miguel. The noisy voter model on complex networks. *Scientific Reports*, 6(1):1–14, 2016.

[72] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss. Threshold-limited spreading in social networks with multiple initiators. *Scientific Reports*, 3:2330, 2013.

[73] R. Milo, N. Kashtan, S. Itzkovitz, M. E. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

[74] J. Blitzstein and P. Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489–522, 2011.

[75] G. F. de Arruda, G. Petri, F. A. Rodrigues, and Y. Moreno. Impact of the distribution of recovery rates on disease spreading in complex networks. *Physical Review Research*, 2(1):013046, 2020.

[76] L. F. Lafuerza and R. Toral. On the effect of heterogeneity in stochastic interacting-particle systems. *Scientific Reports*, 3(1):1–8, 2013.

[77] V. Barros, L. Liao, and J. Rousseau. On the shortest distance between orbits and the longest common substring problem. *Advances in Mathematics*, 344:311–339, 2019.

[78] R. Arratia and M. S. Waterman. An Erdös-Rényi law with shifts. *Advances in Mathematics*, 55(1):13–23, 1985.

[79] M. F. Schilling. The longest run of heads. *The College Mathematics Journal*, 21(3):196–207, 1990.

[80] L. Gordon, M. F. Schilling, and M. S. Waterman. An extreme value theory for long head runs. *Probability Theory and Related Fields*, 72(2):279–287, 1986.

[81] B. Eisenberg. On the expectation of the maximum of IID geometric random variables. *Statistics & Probability Letters*, 78(2):135–143, 2008.

[82] F. Herrerías-Azcué and T. Galla. Consensus and diversity in multistate noisy voter models. *Physical Review E*, 100(2):022304, 2019.

[83] C. Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.

[84] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[85] V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Physical Review E*, 77(4):041121, 2008.

[86] A. F. Peralta, A. Carro, M. San Miguel, and R. Toral. Analytical and numerical study of the non-linear noisy voter model on complex networks. *Chaos*, 28(7):075516, 2018.

[87] A. Nematzadeh, E. Ferrara, A. Flammini, and Y.-Y. Ahn. Optimal network modularity for information diffusion. *Physical Review Letters*, 113(8):088701, 2014.

[88] A. Clauset, E. Tucker and M. Sainz. The Colorado Index of Complex Networks. 2016. https://icon.colorado.edu/, Accessed: 2019-08-05.

[89] D. E. Knuth. *Stanford GraphBase: A platform for combinatorial computing*. Addison-Wesley, 1993.

[90] R. R. Faulkner. *Music on demand: composers and careers in the Hollywood film industry*. Transaction Books, 1983.

[91] S. C. Freeman and L. C. Freeman. *The networkers network: A study of the impact of a new communications medium on sociometric structure*. School of Social Sciences University of Calif., 1979.

[92] S. F. Sampson. *A novitiate in a period of change: An experimental and case study of social relationships.* PhD thesis, Cornell University, 1968.

[93] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha. Eigenvector-based centrality measures for temporal networks. *Multiscale Modeling & Simulation*, 15(1):537–574, 2017.

[94] V. Krebs. Uncloaking terrorist networks. *First Monday*, 7(4), 2002.

[95] R. S. Burt. Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 92(6):1287–1335, 1987.

[96] B. Kapferer. *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*. Manchester University Press, 1972.

[97] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

[98] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):065103, 2003.

Appendix A

Further investigations of the quoter model

A.1 Quoter model parameters

To support our results, here we explore other choices of quoter model parameters (q and λ). The simulations are done on smaller networks to make it less computationally expensive to do a wide sweep of the parameter space. We first simulate the quoter model on ER, BA, and small-world networks for $q \in \{0.1, 0.5, 0.9\}$ and vary $\langle k \rangle$ or the rewiring probability, p, to support results from Sec. 3.2.1 and Sec. 3.2.2. We then simulate the ER, BA, and small-world experiments again for various combinations of the quote probability q and mean quote length λ. We evaluate the robustness of results for ER networks as follows. For each combination of (q, λ), we calculate the difference $\langle h_\times \rangle_{k=20} - \langle h_\times \rangle_{k=6}$, whereby $\langle h_\times \rangle_{k=20}$ we mean the average cross-entropy on ER networks of average degree $k = 20$. The quantity will be positive if density inhibits information flow. This allows us to assess the how the magnitude of our

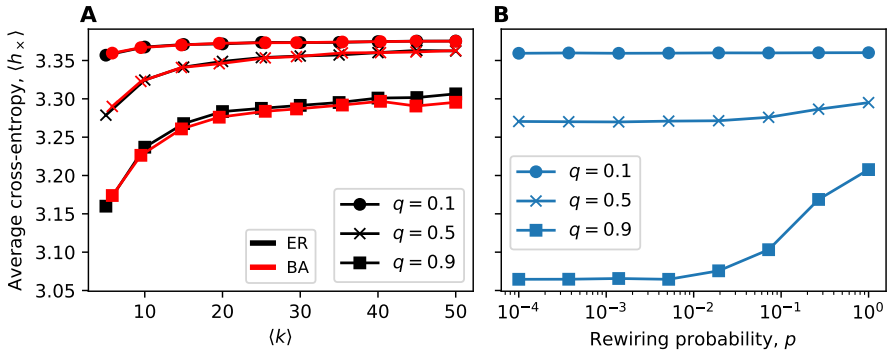

Figure A.1: Trends in information flow in ER, BA, and small-world networks for $q \in \{0.1, 0.5, 0.9\}$. With the exception of very low quote probabilities, we see qualitatively similar trends. ***(A)*** *ER & BA networks of size $N = 100$ with varying average degree. Each point constitutes 200 simulations.* ***(B)*** *Small-world networks of size $N = 200$ with $k = 6$ with varying rewiring probability. Each point constitutes 500 simulations.*

results vary with (q, λ), although it does not confirm a monotonic trend holds.

We repeat these calculations with the BA networks and extend them to the small-world networks by replacing $\langle k \rangle$ with $p \in \{0, 1\}$. In general, we find in Figs. A.1 and A.2 that our results are qualitatively robust to parameter choices, with the exception of very small values of q, as we expect.

A.2 Summarizing the cross-entropy

In this work, we summarized $h_\times$ by the mean $\langle h_\times \rangle$ and variance $\mathrm{Var}(h_\times)$. In Fig. A.3, we see that this choice was appropriate: examining the distributions of $h_\times$ for various networks shows that they are approximately normal. We also find the mean and median $h_\times$ to be approximately equal.

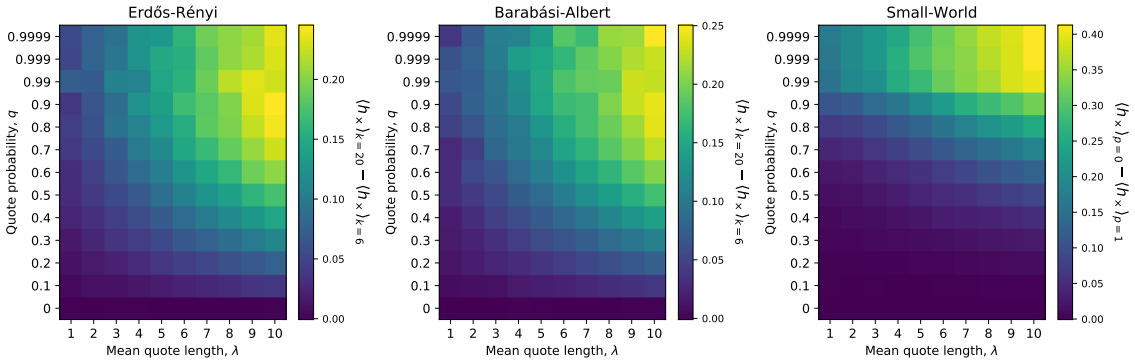

Figure A.2: Effects of quoter model parameter choices on observed trends. Information flow is lower for denser ER and BA networks across a range of q and λ with the effect being more pronounced at higher values of q and λ. Likewise, for small-world networks, more clustering (lower p) exhibits higher $h_\times$ than less clustering (higher p), with the effect being most pronounced at $q > 0.5$ regardless of λ. Here, ER & BA networks had $N = 100$ and small-world networks had $N = 200$ and $k = 6$. Each cell constitutes 100 simulations.

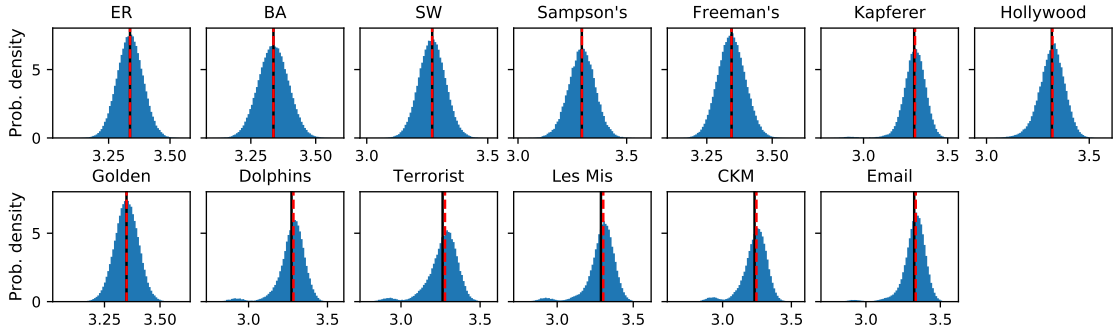

Figure A.3: The distributions of $h_\times$ for quoter model simulations on various networks. Examining the distributions supports using $\langle h_\times \rangle$ and $\mathrm{Var}(h_\times)$ as summary statistics, although some real networks show a small bimodality (an excess of $h_\times < 3$ bits). We also remark that the mean and median are approximately equal (solid line shows $\langle h_\times \rangle$, dashed line shows median $h_\times$) for all networks. ER & BA networks have $N = 1000$ nodes with $\langle k \rangle = 12$, and 200 simulations as in Fig. 3.1. Small-world networks have $N = 200$ nodes with $k = 6$ and $p = 10^{-4}$, and 500 simulations as in Fig. 3.4A. Real-world networks are from 300 simulations as in Fig. 3.2 and Fig. 3.4B,C. Quoter model parameters are given in Sec. 3.1.1.

Appendix B

Network corpus

All networks studied here can be found through the Index of Complex Networks (ICON) [88]. We converted any directed or weighted networks to undirected (bi-directional) and unweighted. Details for each of the ten networks:

1. Les Miserables co-appearances [89] [Undirected, Weighted].
2. Hollywood film music [90] [Undirected, Weighted]. This is a bipartite network; we converted it to a one-mode projection (nodes are composers and two composers are linked if they worked with the same producer).
3. Freeman's EIES dataset [91] [Directed, Weighted]. We used the "personal relationships (time 1)" network.
4. Sampson's monastery [92] [Directed, Weighted]. We used the Pajek dataset. The weight of a directed link represents how an individual rates the other. The rating can be positive (1,2,3 = top 3 ranked) or negative (-1,-2,-3 = worst 3 ranked). We chose to only keep links which were positive.
5. Golden Age of Hollywood [93] [Directed, Weighted]. We used the aggregated network over 1909-2009.
6. 9-11 terrorist network [94] [Undirected, Unweighted].
7. CKM physicians social network [95] (1966) [Directed, Unweighted]. We used "CKM physicians Freeman" networks hosted by Linton Freeman, and chose the

"friend" network (i.e., the third adjacency matrix). We took only the giant component.

8. Kapferer tailor shop [96] (1972) [Undirected, Unweighted]. We used the “Kapferer tailor shop 1” Pajek dataset (kapfts1.dat).

9. Dolphin social network [97] (1994-2001) [Undirected, Unweighted].

10. Email network (Uni. R-V, Spain, 2003) [98] [Directed, Unweighted]. We used the “email-uni-rv-spain-arenas” network.