Network analyses of 4D genome datasets automate detection of community-scale gene structure and plasticity

Benjamin Walker\textsuperscript{1,*}, Dane Taylor\textsuperscript{2,3}, Caitlin Hult\textsuperscript{4,5}, David Adalsteinsson\textsuperscript{1}, and M. Gregory Forest\textsuperscript{1,6}

\textsuperscript{1}Department of Mathematics, University of North Carolina at Chapel Hill
\textsuperscript{2}Department of Mathematics, University at Buffalo, State University of New York
\textsuperscript{3}Computational and Data Enabled Science and Engineering, University at Buffalo, State University of New York
\textsuperscript{4}Department of Microbiology and Immunology, University of Michigan Medical School
\textsuperscript{5}Department of Chemical Engineering, University of Michigan
\textsuperscript{6}Departments of Applied Physical Sciences and Biomedical Engineering, University of North Carolina at Chapel Hill

October 2, 2018

Abstract

Chromosome conformation capture and Hi-C technologies provide gene-gene proximity datasets of stationary cells, revealing chromosome territories, topologically associating domains, and chromosome topology. Imaging of tagged DNA sequences in live cells through the lac operator reporter system provides dynamic datasets of chromosomal loci. Chromosome modeling explores the mechanisms underlying 3D genome structure and dynamics. Here, we automate 4D genome dataset analysis with network-based tools as an alternative to gene-gene proximity statistics and visual structure determination. Temporal network models and community detection algorithms are applied to 4D modeling of G1 in budding yeast with transient crosslinking of 5kb domains in the nucleolus, analyzing datasets from four decades of transient binding timescales. Network tools detect and track transient gene communities (clusters) within the nucleolus, their size, number, persistence time, and frequency of gene exchanges. An optimal, weak binding affinity is revealed that maximizes community-scale plasticity whereby large communities persist, frequently exchanging genes.
1 Introduction

The field of chromosome conformation capture (3C) and its Hi-C generalizations continues to generate tremendous progress in understanding how the genome self-organizes, including compelling evidence that chromosomes occupy territories in the nucleus, form topologically associating domains (TADs), and exhibit fractal topology ([28, 37, 54]). Alternative insights arise from imaging of the positional fluctuations of tagged DNA sequences on specific chromosomes in live cells ([9, 36, 20]) through the lac operator/lac repressor reporter system ([5]). Chromosome modeling provides yet a third perspective, generating simulated 4D datasets of the 3D position time series of chromosomes based on assumptions of the physical and chemical mechanisms responsible for structure and dynamics during phases of the cell cycle (cf. [32, 42, 22, 50, 1, 57, 25, 53, 34, 51, 12, 55, 14, 52, 26, 8, 23]).

The role of structural maintenance of chromosome (SMC) proteins that act on top of the genome to guide chromosome organization through hierarchical folding and loop generation continues to mount (cf. [4], [31], [2], [24], [43]).

Our primary goal in this paper is to automate the quantitative analysis of 4D chromosome datasets, generated either from chromosome modeling or high-resolution imaging, with network-based analyses. In this way, we aim to provide an agnostic (user independent) alternative to distance-based visual determination of structure.

We apply temporal network models and analyses ([40, 38, 39]), including temporal community detection ([33]), to 4D datasets generated by polymer model simulations of all 16 chromosomes during G1 phase of interphase in budding yeast. The model and code are identical to those in ([26]), whereby structural-maintenance-of-chromosome (SMC) proteins transiently crosslink 5kb domains in the nucleolus on Chromosome XII. Automated network-based methods scan the 4D datasets over four decades of SMC-binding kinetic timescales, and robustly detect, track and label transient gene communities (i.e., spatiotemporal clusters of 5kb domains) in the nucleolus, while recording summary statistics on the size, number of, persistence of, and exchanges between communities.

We confirm results in ([26]), which were based on distance-based, visual metrics, and further reveal non-monotone behavior in the network statistics. We find an optimum timescale exists that maximizes community plasticity, whereby persistent gene communities emerge and flexibly exchange chromosome domains over time. Characterized by spatial segregation, compaction, and frequent self-interaction, the nucleolus is a readily identifiable subdomain of the genome responsible for ribosomal RNA (rDNA) synthesis. The nucleolus is composed of repeat DNA sequences, posing significant experimental imaging challenges to discern intra-nucleolar dynamics and structure, and presenting challenges to discern among possible mechanisms governing its spatial and temporal self-organization. Toward this goal, [1] explored Chromosome XII in budding yeast, the largest chromosome that also contains the nucleolus. They observed the position time series of 15 loci distributed every 100 kb along Chromosome XII, providing insight into how and where Chromosome XII resides in the nucleus.
relative to the rest of the yeast genome. Their observations were consistent with computational predictions ([58]) of Rouse models of tethered polymers for all 16 yeast chromosomes during interphase. Recently, the Zimmer group has incorporated Bayesian inference with Rouse polymer computational models to determine chromatin properties consistent with an extensive experimental dataset from imaging in fixed and live cells and two Hi-C studies ([3]). These studies, while advancing experimental and mechanistic understanding of Chromosome XII, do not provide intra-nucleolar insights.

The Bloom lab and collaborators, in pursuit of intra-nucleolar mechanistic organization, has explored the binding kinetics of structural maintenance of chromosome (SMC) proteins through polymer modeling and comparisons with live cell imaging of budding yeast. SMC proteins are known to transiently bind to chromosome domains, and in particular to the nucleolus within Chromosome XII. In [52, 26], the impact of permanent and transient SMC protein-mediated chromosomal crosslinks were simulated, both exclusively within the nucleolus and also externally to allow additional active chromosome domains outside the nucleolus. The spatiotemporal self-organization within the nucleolus and genome-wide were explored during the G1 stage of the cell cycle and the role of SMC proteins was investigated by considering permanent, weak (short-lived), and strong (long-lived) transient crosslinking kinetics. Simulated 4D datasets were passed through the same microscope simulator pipeline as the experimental videos. Among other insights, these studies show that the experimentally observed compaction of the nucleolus into a crescent shape near the nuclear wall during G1 is matched best with weak (short-lived) transient SMC protein-mediated crosslinking kinetics within the nucleolus.

N.B. The above crosslinking mechanism imposes an explicit, indiscriminate function of SMC proteins between any two, non-nearest-neighbor, nucleolar domains on the order of 5 kb along Chromosome XII. There are 361 domains in total within the nucleolus in [26], providing a rich pool of available domains to crosslink. Domains have a probability to bind to one another when they are within a threshold distance set by the size of the SMC binding protein. Once a pair of domains is bound, there is a probability to unbind, based on a presumed SMC protein affinity to the active domains. This mechanism generates massive nucleolar compaction together with 3D, transient, “structure-within-structure” organization that varies with the presumed binding-unbinding kinetics, as shown in [26] and additionally in the present study.

An alternative, discriminatory mechanism is "lengthwise compaction", i.e., nearest neighbor loop extrusion ([2]) through SMC proteins (condensin, cohesin) that operate with topoisomerase, generating loops and compacting individual chromosomes ([24]). These two loop-generating mechanisms presumably can act independently and in tandem. We do not incorporate loop extrusion in the present study as we are building network tools to recapitulate and extend the distance-based results in [26]. We refer the reader to [24] for loop extrusion modeling, to [43] for tandem loop-generating modeling, and references cited by those two studies. Nonetheless, we emphasize that the temporal network analyses and automated algorithms that we present here can be applied to any
4D chromosome dataset, including simulated and experimental data describing the entire genome, a specific chromosome, or segments of chromosomes (and with any sources of SMC organization).

These dynamic chromosome conformation models provide complementary information to chromosome conformation capture (3C) techniques and genome-wide Hi-C techniques (16, 13, 28, 17, 26, 37, 54). 3C and Hi-C experiments rely on population averages of gene-gene proximity on all chromosomes over thousands of dead cells whose chromosomes have been permanently crosslinked by formaldehyde. Many powerful inferences have been drawn from Hi-C data and stochastic polymer modeling, drawn from the analyses of empirical and synthetic datasets encoding maps related to the pairwise distances between genes that are spatially nearby.

In the present study, we explore inferences that can be drawn from spatiotemporal (4D) genome data by network analyses and community detection (35, 33). There is a growing interest to analyze Hi-C datasets and model chromosome interactions using network models (40, 38, 39), which has opened the door to study chromosomal datasets using network-based algorithms including community detection (59, 11, 9, 45) and centrality (29, 30). These network methods have quantified how chromosome conformational changes can precede changes to transcription factors and gene expression (10), leading to new approaches for cellular reprogramming (11, 29). Our network analyses cater specifically to temporal network datasets, complementing previous research (11, 9) identifying topologically associating domains (TADs) in Hi-C data using community detection methods designed for static networks. We apply temporal community detection algorithms including the Louvain algorithm for multilayer modularity (33) to automate the identification of spatially and temporally robust clusters, or communities, of genes from 4D simulation datasets. To our knowledge, network theory has yet to be employed to study 4D genome datasets from Rouse-like polymer models of chromosomes, and we propose network theory as a common mathematical framework bridging the study of synthetic and experimental 4D nucleome data.

We focus on the Rouse-like polymer model of 26, for which clustering has been previously observed. Using distance-based visual metrics, 26 identified several aspects of intra-nucleolar clustering occurring due to transient SMC-protein-mediated crosslinking: (i) “structure within structure” multiscale organization in the presence of weak, short-lived crosslinking kinetics; (ii) the evaporation of clusters as the crosslinks are made stronger and longer-lived; and (iii) all substructure appears to vanish in the limit of strong, long-lived crosslinking. Related experimental-modeling studies by 19 identify persistent subcompartments surrounding the nucleolus associated with phase-separated droplets, each comprised of macromolecular complexes. While these features relate to the nucleolar environment surrounding chromosomal DNA in the nucleolus, our focus here and in 26 is the behavior of the DNA itself, and how robust sub-structures at multiple scales (i.e., pairs of 5 kb domains up to clusters involving dozens of 5 kb domains), are tuned by the timescales of SMC crosslinking kinetics.

The 4D data analyses used in 26 consisted of proximity heat maps of the
mean distance over time between all pairs of beads in the nucleolus, histograms of bead-bead distances at each timestep, 3D snapshots of the spatial organization of the nucleolus, and time evolution of these metrics. Here, we introduce novel network analyses for nucleolar structure and dynamics by developing temporal networks to encode timeseries datasets describing the distances between 5 kb domains. The goal is to define agnostic, user-independent, biologically-informed metrics to identify, extract, quantify, and visualize substructure of the nucleolus within Chromosome XII during the G1 stage of interphase, and to incorporate persistence over time into the cost function identifying substructure – namely, we identify “spatiotemporal gene communities”. We present a dynamic substructure (community) identification approach that does not rely on visual identification of clustering and self-segregation, whether from heat maps or 3D snapshots or 4D movies. We detect spatiotemporal organization beyond two-point statistics, and identify the timescales over which organization persists, linking the timescales to the cluster identification algorithm, and identifying whether clustering occurs on multiple scales. We find that specific tasks (e.g., nucleolus-wide cross communication, marked by segregation into disjoint clusters of highly interacting genes and high gene exchange rates between clusters) are optimized at a "sweet spot" in the short timescale range of binding kinetics.

An important finding in [26] is that closer agreement with live cell imaging arises when the SMC protein-binding kinetics are weak, i.e., the crosslinks are short-lived. The temporal network analyses developed herein provide a network-based interpretation of this phenomenon, and reveal a surprising new aspect: we show that within the short-lived bond duration range, there is a remarkable sensitivity to the crosslinking timescale, marked by a non-monotonic range of optimality for the "cross-communication", or gene-gene mixing, among all active binding domains in the nucleolus. We quantify cross-communication by observing the total number of self-organized communities (domain clusters), the number of gene domains per community (cluster size), and community plasticity, which we characterize by the duration of (and time-between) instances in which 5 kb domains are in the same community (and are said to "communicate"). From a biological perspective, this optimality induced by many weak molecular anchors, is counter-intuitive at first blush. Namely, it suggests that a DNA crosslinking protein family with a high turnover rate is more effective for establishing the heterogeneous self-organized clustering of chromosome domains, leading to higher levels of compaction than a DNA-crosslinking protein family yielding stable, long-lived crosslinks. It further suggests a mechanism through which distal regions of the genome can achieve, and remain in, close proximity for tunable timescales to perform diverse functions (e.g., manufacture of ribosomal RNA, DNA repair, transcription). The optimal weak binding affinity is likewise shown to minimize the waiting times for nucleolus genes to interact in pairs, or communities, influencing the time-dependent aspects of biological functions requiring (potentially) a sequence of chromosome conformations.
2 Model

Chromatin dynamics within confined yeast nuclei have been widely modeled using Rouse polymer bead-spring chain models; see earlier citations. We employ the identical model and code in [26] for interphase, resolving the entire yeast genome into 2803 total beads (chromosome domains or tension blobs), where each bead represents approximately 5k base pairs, and the number of beads on each of the 32 chromosome arms reflects their experimentally identified length. The beads experience Brownian, entropic, repulsive and hydrodynamic drag forces, and are physically confined to the nucleus. Each chromosome arm is tethered at both ends to the nuclear wall: all emanating from the centromere at one end, with the other end tethered to one of six telomeres. Along each arm, entropic nonlinear, worm-like chain springs connect neighboring beads. For each simulation, we implement random initial conditions and simulate approximately 20 minutes of G1 during interphase.

Further, SMC proteins are allowed to transiently crosslink pairs of non-neighboring beads in the nucleolus, discretized as a contiguous chain of 361 beads on chromosome XII. (We note that [26] also studied when the active beads are partitioned on separate chromosomes and observed it to have negligible affect on the multiscale self-organized clustering.) The transient crosslinks are stochastically created between pairs of nucleolar beads; we defer to [26] for the details. The key timescale in the SMC protein binding and unbinding dynamics is $\mu$ (the "mean on" time in [26], measured in seconds), which dictates two states for each individual bead. A bead is active (meaning it will form a transient bond if there is at least one other bead within the threshold distance of $d = 90$ nm) for a time that is a random variable drawn from the normal distribution $N(\mu, (\mu/5)^2)$. Once the active state ends, any transient bond that may have arisen breaks; the inactive state of that bead persists for a new time that is a random variable drawn from $N(\mu/9, (\mu/45)^2)$. Hence, bonds (crosslinks) are established and broken stochastically, and the single parameter $\mu$ thereby dictates the kinetic timescales for binding and unbinding. In [26], three values of $\mu \in \{0.09, 0.9, 9.0\}$ were studied. In this paper, we explore a wide range of $\mu$ values logarithmically spaced between 0.09 and 90 to allow for a more detailed investigation of how nucleolar dynamics and self-organization are affected by $\mu$. See [26] for additional model details.

3 Results

We explore how the kinetic timescale parameter $\mu$ of SMC-protein indiscriminant crosslinking influences the formation and temporal evolution of intranucleolar structure and heterogeneity over 4 decades of timescales for $\mu$. We generate 4D datasets from simulations of the model in [26], and then apply the network models, analyses, and fast algorithms whose description we defer to the Methods section.
Video 1: One minute of real-time simulation for $\mu = 0.09$, demonstrating association of beads into large, stable clusters. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset0_allRed_finer_realtime.mp4.

Video 2: One minute of real-time simulation for $\mu = 0.19$, showing association of beads into clusters that exhibit frequent changes in membership. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset6_allRed_finer_realtime.mp4.
3.1 Transient Binding Timescales Influence Intra-Nucleolar Clustering

We first study various representations of the bead positions taken from simulations with three selected values of the binding parameter $\mu \in \{0.09, 0.19, 1.6\}$. “Snapshots” of the simulated model data are shown in Figure 1(A–C), with beads in the nucleolus on Chromosome XII highlighted in blue and all remaining chromosome beads in gray. In Figure 1(D–F), we show only nucleolar beads, color-labeled by the network community detection analyses that we describe in the following sections. We also show videos of the time evolution of the beads, along with a simulated microscope projection, in Videos 1 to 3 for each of the three chosen timescales. Based on Figure 1(A–C) and Videos 1 to 3, we identify three qualitative regimes for intra-nucleolar clustering:

1. **rigid clustering**: strong, stable clusters for very short-lived bond durations such as prescribed by $\mu = 0.09$ (D).

2. **flexible clustering**: slightly weaker clustering whereby communities persist however beads flexibly enter and leave communities, which occurs, for example, when $\mu = 0.19$ (E).

3. **non-clustering**: a lack of robust communities in which beads act as lone units or in bead pairs in the "slow" timescale, e.g., $\mu = 1.6$ (F).

We will continue to use this terminology when referring to these three timescale regimes.
We note that the dissolution of the visible sub-structure reported in [26] at \( \mu = 90 \), detected here as the dissolution of robust community structure, has already occurred at \( \mu = 1.6 \).

In Figure 1(G–I), we show the bead-bead distances associated with the bead positions in (D–F) (which are identical to those in (A–C)). We depict the bead-bead distances using heatmaps as described in Methods Section High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps. Such heatmaps are widely used to depict pairwise-count data from Hi-C ([18] [44] [21] [4] [27] [28]) and simulation data from 4D polymer models ([53] [52] [26]). We point out that panels (G–I) depict heatmaps of bead-bead distances at a single instance in time for a single cell, giving very precise information. Heatmaps arising from Hi-C data are constructed from population averages, and heatmaps for 4D simulated datasets from models may also reflect time-averaged distances across simulation durations. This averaging may cause information to be obscured from the heatmaps. Heatmaps have been used to identify sub-structures among genes, or clusters of genes, in the above Hi-C and modeling references. Comparing the second and third column of Figure 1, we note the difficulty (false negatives and false positives) in detecting the presence of structure-within-structure by visually examining heatmaps.

In Figure 1(J–L), we reorder the beads according to a standard hierarchical clustering dendrogram algorithm. See Section Naive Cluster Detection for Pairwise Distance Maps for further discussion. This relabeling permutes the rows and columns of heatmaps so that nearby rows/columns correspond represent nearby beads in the simulation, leading to the formation of dark blocks along a heatmap’s diagonal. While these blocks are clearly visible in the fast crosslinking case, and slightly less visible but still present in the intermediate crosslinking, they also appear with the slow crosslinking even though we know there is no clustering (see panel F). Our conclusion is that there is a need for a more reliable and objective method to study the clustering of chromosome domains in the nucleolus, especially spatiotemporal methods that take into account how bead positions change with time (as opposed to Figure 1, which describes the beads’ positions at a single instance in time). Herein, we will develop temporal network modeling and temporal community detection as a scalable and automated approach to robustly identify such spatiotemporal clustering and characterize the clustering dynamics.

### 3.2 Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale

To further motivate our development of network theory for chromosome models, we develop here a spatiotemporal two-point analysis of the 4D data, using alternative statistics to gene-gene proximity heat maps. This analysis reveals that the binding parameter \( \mu \) strongly influences gene-gene interactions at a local level. (We further note that these distance-based metrics will be generalized.

---

[^1]: https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.cluster.hierarchy.dendrogram.html
We first define gene interactions: events in which the distance between any pair of beads drops below a critical distance $d^*$ . All pairs of beads that are nearer than distance $d^*$ are said to be interacting. Gene mixing refers to the frequency and total count of gene interactions. We quantitatively study gene mixing through the following statistics:

(A) The interaction fraction indicates the fraction of possible unique bead pairs that interact at least once during an interphase simulation.

(B) The mean interaction number indicates the number of simultaneous interactions (i.e., beads within distance $d^*$ ) for a nucleolus bead, averaged across time and across beads.

(C) The mean waiting time indicates for any two beads, selected at random, the average time that passes between their $i$-th and $(i+1)$-th interactions.

(D) The mean interaction duration indicates the amount of time beads enter and reside within the interaction distance.

These statistics depend, of course, on the interaction distance $d^*$ . We surmise that the interactions of beads are correlated with clustering behavior, and so we pick a value of $d^*$ that reflects the spatial scale of the clustering. By observation (see ahead in Figure 3), a reasonable such choice is $d^* = 100$ nm.

In Figure 2(A–D), we study these summary statistics across a 4-decade range of the binding parameter $\mu$ . These metrics reveal three regimes, optimizing different attributes, based on the value of the binding parameter $\mu$ :

The fastest crosslinking ($\mu \lesssim 0.1$ ) yields a self-organized structure that maximizes the number and the duration of gene-gene interactions (see panels B and D for $\mu \approx 0.1$ ). Recall from Video 1 that this $\mu$ value yields many large, stable clusters. This is reflected in the high number of simultaneous interactions (many beads in the same cluster) and low number of unique interactions (beads are restricted to interacting with other beads in their same cluster).

At slightly longer-lived bond duration, $0.19 \lesssim \mu \lesssim 0.9$ , we see flexible clustering behavior from Video 2. Notably, we find here that this flexible clustering has interesting properties beyond simply being a weaker version of the strong clustering from the rigid clustering regime. Namely, Figure 2A shows that these $\mu$ values maximize the fraction of unique pairs of beads that interact at least once over the simulation, and Figure 2C shows that these values minimize the waiting times between subsequent interactions. Thus, we can say that the flexible clustering regime promotes global pairwise interactions among all active beads. SMC proteins with such an intermediate crosslinking timescale will thereby promote collective interactions among all active domains. These circumstances would accelerate a homology search, for example, to facilitate DNA repair, if the sister chromosomes were suddenly activated by a family of SMC proteins whose binding affinity was near this “sweet spot”.
The $\mu \gtrsim 1$ is associated with a non-clustering regime, as shown in Video 3. The lack of clustering is reflected in the low number of simultaneous interactions, and the freely diffusing nature of the beads is reflected in the low interaction duration and relatively high interaction fraction, but not as high as in the flexible clustering regime.

3.3 Histograms of 2-Point Pairwise Distances between Nucleolar Beads

Our network analyses rely on the distances between bead pairs, and so we first study these distances and as well as the effects of averaging the distances across a time window or across multiple cells (e.g., multiple simulations). These experiments illustrate how clustering — which inherently describes multiway relationships — can be studied through pairwise distances — which inherently describe two-way relationships —, and how there remain important open problems related to the timeseries signal processing of 4D chromosome conformation datasets. We describe our computation of such distances in the Methods Section High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps.

In Figure 3(A), we plot the distribution of all pairwise distances $\{d_{ij}(t)\}$ at a single time $t$ for three kinetic timescales, given by the same values $\mu = 0.09, 0.19, 1.6$ as shown in Figure 1. For $\mu = 0.09$, the pairwise distance distribution is clearly a multimodal distribution ([26]). The peak near $d \approx 50$ represents a large number of very short pairwise distances between beads in the same cluster. For slightly larger $d$, the density drops to zeros, indicating a separation distance between clusters. Interestingly, we observe two more peaks near $d \approx 300$ and $d \approx 600$. The clarity of these peaks suggests that the clusters themselves are regularly spaced from one another, reminiscent of a lattice structure. This shows the three layers of the multiscale structure of the nucleolus for $\mu = 0.09$: its existence as a dense, secluded section of the nucleus, the self-organization of intra-nucleolar clusters, and the individual beads within each cluster. For $\mu = 0.19$, one can also observe in Figure 3(A) three peaks in the empirical probability density for bead-bead distances, but these peaks are much less pronounced. This shows a gradual transition in the degree of clustering as we increase $\mu$. There is also a smaller gap between peaks. Together, these observations recapitulate our observations in Figure 1(E), wherein the clusters can be observed to be less compact. Finally, for $\mu = 1.6$, there is no multimodal structure in the bead-bead distance plot. This is consistent with our expectation that there is no clustering structure present for this range of $\mu$.

The rigid and flexible clustering cases differ not only in how strong the clustering is at any given time, but also in how stable the structure is in time. We investigate this by considering how averaging pairwise-distances either across across time (Figure 3(B)) or over multiple simulations (Figure 3(C)). influences pairwise-distance probability densities.

In Figure 3(B), we plot the empirical probability densities for pairwise dis-
stances averaged across our 20 minute simulations. Note for $\mu = 0.19$ that the density is no longer multimodal, implying that aggregating the data across a large time range inhibits the detection of flexible clusters, which by definition change with time. Note that the rigid clusters, which are very stable across time, remain discernable as the pairwise probability density remains multimodal. Unsurprisingly, the slow crosslinking appears qualitatively very similar in the long time average, as there was no apparent structure in the first place.

In Figure 3(C), we plot the empirical probability densities for pairwise distances at a single time but averaged across 10 simulations with different random initial conditions. Note for all $\mu$ that there is no longer any multimodal structure for these densities, highlighting that averaging across heterogeneous cell populations obscures the detection of clusters.

### 3.4 Automated Interaction Analysis with Network Models and Network Algorithms

We now reformulate the 4D datasets using temporal network modeling. We first obtain a collection of bead-bead distances as described in Methods Section High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps. Given a set of pairwise distances, we construct a network so that each bead is represented by a node, and an edge is made between any two beads whose distance is less than a chosen threshold $d^*$ (analogous to the gene interactions described in Section Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale). In order to preserve information about the distance, these edges are assigned a weight based on an exponentially decreasing function of distance. We refer to such networks as gene-interaction networks and we provide further details in the Methods Section Gene-Interaction Networks for Rouse-like Polymer Models.

First, we point out that the four summary statistics (A)–(D) from Section Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale can be interpreted as network properties: interaction fraction equals the fraction of unique pairs of nodes that share an edge at least once during the simulation; mean interaction number equals the time-averaged unweighted node degrees; mean waiting time equals the average time that passes between subsequent edges for a fixed pair of nodes; and mean interaction duration indicates the average amount of time in which an edge continually persists between two nodes. Thus, the study in Section Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale can be interpreted as an analysis of gene-interaction networks focused specifically on local network properties—particularly, the diversity, number, duration, and wait-time between edges in the gene-interaction network.

Figure 4(A–C) illustrates gene-interaction networks constructed for $\mu = 0.19$ (see the second row in Figure 1) for three choices of distance threshold $d^*$. Note how additional edges appear as $d^*$ increases, leading to a more connected network. In particular, increasing $d^*$ decreases the number of connected components, defined as a set of nodes and edges such that there exists a path between
any two nodes in a given connected component. Connected components can be used to identify gene clusters, and studying the number of connected components as a function of $d^*$ is a multiscale approach that provides insight into the spatial scale of clustering. In the case of very dense, uniformly well-separated clusters, there should be a wide range of threshold values for which the connected components are invariant when the threshold sweeps the range of distances between beads in different clusters.

To confirm this, we plot the number of connected components versus $d^*$ in Figure 4(D–F). Additionally, we overlay the same pairwise distance density shown in Figure 3(A,B) to show the relationship between presence of edges and decrease in connected components. This is shown in Figure 4(D), showing that there is indeed a very wide plateau in the number of connected components, which corresponds to the range of distance values for which there are very few connections. This shows a network perspective of the observation we made earlier.

In contrast, the same plateau is not present in the intermediate and slow crosslinking panels (E,F respectively). The lack of a plateau for $\mu = 0.19$ highlights the drawbacks of using connected components for cluster analysis; it requires large spatial separation between clusters and is therefore not a robust measurement for clustering. Moreover, it neither takes into account the edge weights that encode distances nor the time-varying aspect of communities. We introduce more advanced community (i.e., cluster) detection techniques in the following sections.

3.5 Robust Multiscale Gene-Community Detection for Single-Layer Networks

Before presenting our study of time-evolving, sub-nucleolar, gene clusters and their impact on gene-gene mixing, we first discuss a method for cluster detection based on network community detection for nontemporal data—specifically, we study a single-layer network representing a set of pairwise distances at a single timestep.

In the Methods Section Multiscale Community Detection for Single-Layer Networks, we describe a community-detection algorithm that we use to algorithmically identify communities (i.e., clusters) of genes: the Louvain modularity-optimization algorithm [7, 33]. Modularity optimization is a multi-resolution method that uses a resolution parameter $\gamma$ to tune the size of found communities. We expect that when a community-detection algorithm finds strong communities, the communities detected should be insensitive to parameter choices. We can determine robustness by running the algorithm for a range of values of $\gamma$ and observing properties of the detected communities. In Figure 5(D–F), we show the how the resolution parameter affects the number of communities for each of the three $\mu$ values. We see that there is a wide plateau of stability in panel (D), corresponding to the rigid clustering regime, for $\gamma \in (3, 10)$. In this region, the modularity algorithm produces a community partition very similar to the CC analysis, as indicated by the high adjusted mutual informa-
tion (AMI), and the number of communities stays stable. For $\gamma \geq 10$, the number of communities rapidly increases as the algorithm starts to divide the communities/clusters into isolate nodes/beads.

In the flexible clustering regime, we observe a region where the slope is slightly shallower but does not exhibit the same stability as that shown in panel (A). This indicates that the algorithm does not find communities that are as robust, consistent with our understanding that the clustering is weaker. The high AMI near $\gamma \approx 5$ in panel (E) shows significant agreement between the clusters identified by the community detection algorithm and CC analysis.

For the non-clustering regime, there is a steady increase in the number of communities with respect to the resolution parameter, which is consistent with our understanding that there is not a scale at which meaningful clustering exists.

Thus we have shown that in the case of very clear clustering, as shown in the fastest crosslinking case, i.e., very short bond duration timescale, we identify very robust communities in the gene-interaction network. This robustness allows us to distinguish between legitimate clusters and noise (c.f. Figure 1(G–L)).

Importantly, the results in Figure 5 focus on position data at a single instance in time, and neglect the fact that clusters change with time. The next step is to introduce an algorithm that explicitly places value (in the cost function) on the temporal information present in the 4D dataset to improve detection of transient, poorly-separated clusters.

3.6 Formation and Dynamics of Intra-nucleolar Gene Communities

Here, we classify and label the self-organized intra-nucleolar clusters using a leading community-detection algorithm for temporal networks. Given a time-varying gene-interaction network (see Methods Section Gene-Interaction Networks for Rouse-like Polymer Models), we apply the Louvain algorithm optimizing multilayer modularity [33]. See the Methods Section Spatiotemporal Cluster Analysis via Community Detection for Temporal Networks for algorithm details. We identify meaningful parameters for the community-detection algorithm using a variety of methods as described in the Methods Section by Spatial-Scale Calibration with Connected-Component Analyses and Calibrating Multilayer Modularity with Layer-Matching Single-Layer Modularity.

In Videos (4–6), we illustrate the effectiveness of the automated algorithm for identifying and labeling spatiotemporal clusters. Videos (4–6) are identical to Videos (1–3), except that the 5 kb domains are colored according to their community label. One can observe that each network community, which is a set of nodes/domains, persists over a period of time—its “lifetime”—and that the nodes/domains can potentially enter and exit the communities. Temporal communities and spatiotemporal clusters are fundamentally different than spatial clusters and single-layer communities since during the existence of a community, its nodes at one instance in time can completely differ from that at another instance in time. That is, analogous to Theseus’s paradox (where it is questioned if a ship in which all wood planks have been replaced over time is still the same
Video 4: One minute of real-time simulation for the rigid clustering regime with $\mu = 0.09$. We observe that the stable and well-separated clusters have been distinctly labeled by the community detection algorithm. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset0_color_finer_realtime_altView.mp4.

Video 5: One minute of real-time simulation for the flexible clustering regime with $\mu = 0.19$. Here the clusters are not so clearly separated, but the colored labels still appear consistent with what one would expect. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset6_color_finer_realtime_altView.mp4.

ship), temporal communities and the spatiotemporal clusters they represent obtain an identity beyond just the set of beads from which they are composed at some point in time.

We define the lifetime of each community as the amount of time that passes between the first and last timesteps for which a bead is present in that community. By construction, at least one or more beads must be in a community
Video 6: One minute of real-time simulation for the non-clustering regime with $\mu = 1.6$. Here there are no clusters present in the data, but the community detection algorithm still tries to give the same label to nearby beads. Due to the lack of stable community structure, beads change label more frequently than for smaller values of $\mu$. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset12_color_finer_realtime_altView.mp4 through the duration of its lifetime.

Using temporal community detection, we are now finally able to quantitatively support our first observations made in Section Transient Binding Timescales Influence Intra-Nucleolar Clustering that there are three distinct clustering phases: a rigid clustering at the shortest binding kinetics timescales, a flexible clustering at slightly slower timescales, and no clustering at all at slower yet timescales. We support these observations by studying the properties of detected communities. In Figure 6(A–C), we plot the average lifetime of communities as a function of their average size (averaged over time). Panels (A–C) indicate the three clustering regimes with $\mu \in \{0.09, 0.19, 1.6\}$, respectively. For the rigid clustering regime, in Figure 6A, we see that most of the communities are large, with an average size of 10 or more beads, and also have a long lifetime of over 100 seconds. This is consistent with our prior observations (e.g. Video 1) that showed large clusters that appeared very stable in time. We also see that the large clusters survive for much longer than the small clusters.

For the flexible clustering regime, in Figure 6B, we can see the same general trend that larger communities tend to have a longer lifetime than smaller communities, but there is a much wider spread of cluster sizes, with only a moderate number of large clusters.

For the non-clustering regime, there appears to be little relationship between cluster size and stability beyond an average size of approximately 3 beads. The clusters also tend to be much smaller, with almost no clusters with an average size over 10 beads.
In Figure 6(D–E), we plot the probability that a bead remains in the same community upon the next timestep, again as a function of cluster size. Panels (D–E) indicate results for $\mu \in \{0.09, 0.19, 1.6\}$, respectively. In agreement with panels (A–C), one can observe that larger clusters are more stable. Note also that the communities exhibit more plasticity for $\mu = 0.19$ than for $\mu = 0.09$ since beads have a higher average probability for changing the community to which they belong.

### 3.7 Community-Level Cross Communication Determines Gene-Level Mixing

We start with a definition. All chromosome domains in the same community are said to be communicating, and we use cross communication to refer to exchanges of domains between communities. Importantly, unlike our study in Section Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale focusing on the dynamics of 2-point relationships related to pairwise distances and/or pairwise interactions, cross communication describes the dynamics of the multi-point relationships induced by the beads’ associations with time-varying (temporal) communities. Cross communication can be thought of as a coarse-grained, community-level representation of pairwise gene-gene mixing.

We proceed by defining summary statistics for cross communication analogous to the 2-point summary statistics for pairwise gene interactions that we previously defined in Section Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale.

(A) The mixing fraction indicates the fraction of bead pairs that are in the same community at least once during a simulation.

(B) The average beads in community indicates, for a nucleolus bead, the average number of beads in the same community at the same time, averaged across time and across beads.

(C) The mean waiting time indicates for any two beads, selected at random, the average time that passes between when they are no longer in the same community and when they are next in the same community.

(D) The mean interaction duration indicates the amount of time between beads when they are first in the same community and when they are no longer in the same community.

In Figure 7, we present summary statistics for cross communication that are analogous to our results in Figure 2 for pairwise gene interactions. Note that the results in Figure 7 are qualitatively identical to those in Figure 2 supporting our hypothesis that gene-level mixing is determined by community-level cross communication. That is, the formation of gene clusters and exchanges of genes between them governs the timescale at which nucleolar domains come in close proximity of one another.
4 Discussion

The dynamic self-organization of the eukaryote genome is fundamental to understanding of life at the cellular level. The last quarter century has witnessed remarkable technological advances that provide massive datasets of both the spatial conformation of chromosomal DNA from cell populations (3C and Hi-C generalizations) and the dynamic motion of domains in living cells (GFP tagging and tracking of specific DNA sequences), from the yeast genome to the human genome. Data mining of this massive data has likewise witnessed remarkable advances in understanding the hierarchical packaging mechanisms of DNA that act on top of the genome, e.g., histones and structural maintenance of chromosome (SMC) proteins, the topology of individual chromosome fibers, their associating domains, and the territories they occupy in the nucleus. The third wave of advances has come from 4D modeling of chromosomes based on stochastic models of entropic, confined polymers, and the coupling of SMC proteins that either bind and crosslink genes on chromosomes or generate loops on individual chromosomes. As these three approaches continue to mature and inform one another, at an ever-increasing pace, insights into the structure and dynamics of the genome continue to deepen.

The motivation for this paper lies in the information that can be inferred from the massive datasets from Hi-C and live cell imaging experiments and from polymer physics modeling. We decided to use the array of tools from network analysis, modeling, and associated fast algorithms to automate the search for dynamically robust structure in the 4D datasets generated by polymer modeling. We note a similar network analysis approach has been applied to Hi-C datasets [40, 38, 39], whereas our datasets have the added feature of temporal information with a very high rate of sampling. Our aim was to infer organization beyond 2-point, time-averaged or population-averaged, gene-gene proximity statistics and heat maps generated from the statistics, and to remove the bias of an individual’s visual determination of structure. To do so, we used the advances in network-based models, data analysis, and algorithms, and applied this arsenal of tools to 4D datasets to: (i) robustly identify clusters, or communities, of genes (5 kb domains, or beads, in our model); (ii) determine the size distribution (number of genes) in such structures; and (iii) determine the persistence lifetimes of communities. In this way, network algorithms automate gene-community detection and persistence, with robustness built in by enforcing insensitivity to algorithm tuning parameters.

We elected to build and implement these network tools on the 4D datasets generated in house, from our recent polymer modeling of interphase budding yeast ([26]). In this model, a pool of SMC proteins transiently and indiscriminantly crosslinks 5 kb domains within the nucleolus on Chromosome XII, and we previously showed that weak, very short-lived binding kinetics provided closer agreement with experimental results (highest degree of compaction of the nucleolus into a crescent shape against the nuclear wall) by the Bloom lab. It was also shown via visualization of the 4D data that weak binding corresponded to a decomposition of the nucleolus into a large number of clusters each consisting
of many 5 kb domains, and these clusters were persistent over time. On the other hand, with strong, long-lived binding, the clusters disappeared.

For the present paper, the sample set of binding kinetics in [26] was expanded to 4D datasets of interphase, sampling over four decades of binding duration timescales. We applied standard, distance-based, 2-point statistical metrics and visualization tools, and then analyzed the full range of 4D datasets with the fast, automated network models and tools. The network algorithms search for and detect time-varying communities (clusters, sub-structures) having spatial and temporal scales that are identified as being robust to algorithm’s multiscale tuning parameters. We then use this information to label and color-code communities, using community-level description to understand the persistence and gene exchanges between communities.

With the above community-scale information and statistics, we generalize standard gene-gene interaction statistics across the four decades of bond duration timescale. As a complement to waiting times for 2 distant beads (5 kb domains) to come within a specific distance of one another, we calculate waiting times for genes to share the same community, and calculate the fraction of all genes that were in the same community at least once during interphase, which we call the community mixing fraction.

From these analyses, we discovered a novel dynamic self-organization regime, wherein the rigid, persistent communities at very short bond duration timescales transition to more mobile (literally, the clusters diffuse faster) communities that interact and exchange genes far more frequently, which we refer to as flexible community structure with enhanced cross-communication. Furthermore, we discovered non-monotonicity in the dynamic self-organization behavior: the community mixing fraction is optimized, coincident with a minimum waiting time for genes to share the same community, at a specific, short-lived but not too short, bond duration timescale.

We emphasize that these network tools and fast algorithms are amenable to any 4D dataset from polymer models. While we restricted the analysis in this study to the nucleolus during G1 of budding yeast where SMC proteins are allowed to transiently crosslink 5 kb domains, the same analyses can be applied to data with tandem SMC crosslinking and loop generation, for any cell type and for any phase of the cell cycle. Moreover, given the growing interest in network-based analyses for Hi-C data, network modeling is well positioned to provide a fruitful direction for data assimilation efforts aimed at connecting simulated and empirical 4D chromosome conformation data. An important challenge facing this pursuit is the development of improved data preprocessing and community-detection methodology for temporal and multimodal network datasets ([45, 47]).
5 Materials and Methods

5.1 High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps

Each simulation of the Rouse-like polymer model yields time-series data \( \{x_i(t)\} \in \mathbb{R}^3 \) that defines the 3D location of each bead \( i \in \{1, \ldots, N\} \) at each discrete time step \( t = 0, 1, \ldots \). We establish a connection between our simulated data and the state-of-the-art in chromosome imaging—namely, high-throughput conformation capture (Hi-C)—by constructing and analyzing pairwise-distance maps. Hi-C “images” the conformation of chromosomes using a combination of proximity-based ligation and massively parallel sequencing, which yields a map that is correlated with the pairwise distances between gene segments. While the actual pairwise distances between gene segments cannot be directly measured, Hi-C implements spatially constrained ligation followed by a locus-specific polymerase chain reaction to obtain pairwise count maps that are correlated with spatial proximity: the count between two gene segments monotonically decreases as the physical 3-dimensional distance between them increases.

To provide an analogue to Hi-C imaging, we construct pairwise-distance maps for our simulated data \( \{x_i(t)\} \). Let

\[
F : \{x_i\}_{i=1}^N \rightarrow \mathbb{R}^{N \times N}
\]

(1)

define a map (used here in the mathematical sense) from a set of \( N \) points \( \{x_i\} \in \mathbb{R}^3 \) to a matrix such that each entry \((i, j)\) in the matrix gives the distance between point \( i \) and point \( j \). Whereas Hi-C imaging aims to study the positioning of chromosomes using noisy measurements that are inversely correlated with pairwise distances, for our simulations we have access to the complete information about the chromosome positioning. We therefore define and study several variations for pairwise distance maps, which will allow us to also study artifacts that can arise under different preprocessing techniques, such as averaging the time series data across time windows and/or averaging across multiple simulations with different initial conditions. We define the following pairwise distance maps:

- An **instantaneous** pairwise distance map \( X(t) = F(\{x_i(t)\}) \) encodes pairwise distances between beads (i.e., chromosome domains) at a particular timestep \( t \).
- A **time-averaged** pairwise distance map \( Y(\tau) = \frac{1}{|\tau|} \sum_{t \in \tau} X(t) \) encodes the pairwise distance between beads averaged across a set of timesteps \( \tau \).
- A **population-averaged** pairwise distance map \( Z(t) = \langle X(t) \rangle_p \) encodes the pairwise distance at timestep \( t \) between beads, which are averaged across several simulations that have different initial conditions (which are chosen uniformly at random).

These pairwise distance maps represent the data that is sought after, but cannot be directly measured, by Hi-C imaging. Moreover, by defining several distance...
maps we are able to study “averaging” artifacts that can arise due to various limitations of Hi-C imaging. For example, Hi-C imaging obtains measurements that are typically averaged across a large heterogeneous distribution of cells that are subjected to nonidentical conditions and exist at nonidentical states in their cell cycles.

5.2 Naive Cluster Detection for Pairwise Distance Maps
Let \( X \in \mathbb{R}^{N \times N} \) denote a pairwise distance map encoding the conformation of chromosomes. The matrix entries \( \{X_{ij}\} \) encode the distances between beads (i.e., chromosome domains), and a large number of clustering algorithms can be applied directly to this type of data structure. All clustering techniques have unique benefits and drawbacks, and the naive application of any algorithm can be problematic. That is, it is generally important to consider numerous complexities for clusterings including the following (potential) complications: In addition to finding clusters, can the algorithm quantify the prevalence of clustering; can it detect if there is a lack of clustering; can it detect clustering phenomena at multiple scales; and can it characterize the time-varying aspects of clusters.

Herein, we analyze clustering with these complexities in mind using advanced community-detection techniques for temporal networks ([33]). It is instructive, however, to first consider the naive application of a popular clustering algorithm that does not carefully consider these complexities. Specifically, we consider a hierarchical clustering algorithm in which points are iteratively aggregated into fewer and fewer groups until all points are aggregated into a single group. According to this aggregation rule, at each step, we merge the two groups that are in closest proximity. We can then order the beads such that any two groups that are merged in the hierarchical clustering appear adjacently in the ordering. This ordering will always emphasize the appearance of clustering, even when no clustering is present, and hence the naive application of such a clustering algorithm falls short of providing a meaningful, robust understanding of clustering. Finally, it’s also worth noting that this clustering algorithm (and more generally, any algorithm that uses pairwise distances \( \{X_{ij}\} \)) cannot be directly applied to Hi-C data, which yields pairwise count data that is related to, but notably different from, pairwise distance data.

5.3 Gene-Interaction Networks for Rouse-like Polymer Models
Given a pairwise distance map \( X \in \mathbb{R}^{N \times N} \) in which each entry \( X_{ij} \) gives the (possibly averaged) Euclidean distance between beads \( i \) and \( j \) as described in 
High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps, we construct a network model in which there are weighted edges (i.e., interactions) only between beads that are in close proximity to each other and for which each edge weight \( A_{ij} \geq 0 \) decreases monotonically with distance \( X_{ij} \). We propose a model with two parameters, \( d^* \) and \( s \), which represent a
distance threshold and a decay rate, respectively. In particular, we define a network adjacency matrix $A$ having entries

$$A_{ij} = \begin{cases} \ e^{-sX_{ij}}, & \text{if } X_{ij} < d^* \\ 0, & \text{if } X_{ij} \geq d^* \end{cases}$$

Note that there exists an undirected edge between $i$ and $j$ (i.e., $A_{ij} > 0$) only when $X_{ij} < d^*$, and $s$ controls the rate in which the edge weight $A_{ij}$ decreases with increasing distance $X_{ij}$. Equation (2) defines a map between a distance matrix and an affinity matrix that encodes a network. Note that for any such adjacency matrix $A$, we can equivalently define the network using the graph-theoretic formulation $G(V, E)$. Here $V = \{1, \ldots, N\}$ denotes the set of nodes (i.e., the beads in the chromosome model) and $E = \{(i, j, A_{ij}) : A_{ij} > 0\}$ denotes the set of weighted edges (i.e., a set encoding which beads are interacting as well as their interaction strengths).

In High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps we defined several versions of pairwise distance maps—instantaneous, time-averaged, and population-averaged maps—and a network model can be constructed for any of these maps:

- An **instantaneous interaction network** refers to a network associated with an instantaneous pairwise distance map $X(t)$.

- A **time-averaged interaction network** refers to a network associated with a time-averaged pairwise distance map $Y(\tau)$. We point out that due to the nonlinearity of (2), a network associated with a time-averaged distance map can in general differ from a temporal average of instantaneous interaction networks, averaged across the same time interval.

- A **population-averaged interaction network** refers to a network associated with an population-averaged distance map $\langle Z(t) \rangle_p$.

In addition to the above network models, we are particularly interested in constructing and studying **temporal interaction networks**, which we define as a sequence of time-averaged interaction networks. In particular, given a sequence of timesteps $t \in \{0, 1, 2, \ldots, T\}$, we partition time into a sequence of time windows $\tau(s) = \{(s - 1)\Delta + 1, \ldots, s\Delta\}$ for $s = 1, 2, \ldots$ of width $\Delta$. We then define a sequence of time-averaged networks $\{G(s)\}$ for $s = 1, 2, \ldots$ associated with these distance maps, which are time-averaged across the nonoverlapping time windows $\{\tau(s)\}$. In practice, we choose the time window width $\Delta$ to be similar to—but slightly larger than—$\mu$. Matching the time-scales of $\mu$ and $\Delta$ allows the temporal interaction networks to efficiently capture the dynamics of interactions. Specifically, if $\Delta$ is too short then the temporal network will be identical across many time steps, which is not an efficient use of computer memory. Moreover, if $\Delta$ is too large, then the temporal network data will be too coarse to identify interaction dynamics occurring at a faster time scale.
5.4 Spatial-Scale Calibration with Connected-Component Analyses

The single-layer modularity has a resolution parameter $\gamma$ that controls the spatial scale of clustering detected. In order to most accurately find the clusters we want, the value of this parameter must be appropriately calibrated based on the spatial scales present in the data. A straightforward way to do this is to study the connected components for comparison. In the fast crosslinking regime, when we know that they are accurate. This is the analysis that is shown in Figure 5.

When performing this analysis, we look for a wide region in the parameter space in which the modularity approach agrees with the connected components approach, indicating that the modularity algorithm is robustly finding the clusters that we expect. By using this calibration technique, we can use the knowledge we have about the easy problem, fast crosslinking, to gain insight into how to best approach the more difficult, slower crosslinking 4D datasets.

5.5 Multiscale Community Detection for Single-Layer Networks

The instantaneous, time-averaged and population-averaged interaction networks are all a type of network commonly referred to as single-layer networks, implying that each is just a single network. This is in contrast with our formulation for a temporal interaction network, which is a sequence of networks. Here we describe the utilization of multiscale community detection algorithms to study clustering (communities) for the single-layer interaction networks. We focus on modularity-based community detection and the Louvain algorithm for modularity optimization since these methods have been widely used in the network science literature and have a well-developed theory base. Moreover, this approach also naturally extends to temporal networks (and multilayer networks, more generally) and represents one of the leading mathematical frameworks to analyze temporal and multimodal datasets.

Modularity-based community detection aims to partition a network into disjoint communities to optimize a modularity measure

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$  \hspace{1cm} (3)

Here, $A$ is the network adjacency matrix (which encodes an instantaneous, time-averaged or population-averaged interaction network), parameter $\gamma$ is a tunable “resolution parameter” allowing one to control the sizes of communities that are detected, $k_i = \sum_j A_{ij}$ is the weighted node degree for each node $i$, $2m = \sum_{ij} A_{ij}$ is twice the total number of edges, $\delta(m, n)$ is a Dirac delta function, and $\{c_i\}$ are integer indices that indicate the community for each node $i$.

The aim of the nonconvex optimization problem $\arg\max_{\{c_j\}} Q$ is to identify for fixed $\gamma$, the optimal community labels $\{c_i\}$ that yield the maximum modularity. We obtain approximate solutions using a popular heuristic widely known
as the Louvain algorithm, which is a greedy iterative method. It begins by placing each node in a separate community, then loops over each node, joining it to an adjacent community if the merger increases $Q$. This continues until there are no remaining mergers that would increase $Q$, yielding near-optimal community indices $\{c_i\}$. The optimal community sizes are dependent on the choice for resolution parameter $\gamma$, and by exploring optimal communities across a parameter range $\gamma \in [\gamma_{\text{min}}, \gamma_{\text{max}}]$, this approach identifies multiscale (and potentially hierarchical) community structure.

5.6 Spatiotemporal Cluster Analysis via Community Detection for Temporal Networks

We analyze spatiotemporal clustering of chromosomes using community detection methodology for temporal interaction networks, particularly an approach based on multilayer-modularity optimization [39]. We study a multilayer modularity measure having the form

$$Q = \frac{1}{2\mu} \sum_{i,j,s,r} \left[ \left(A^s_{ij} - \gamma \frac{k^s_i k^s_j}{2m_s} \right) \delta(s,r) + \omega \delta(i,j) C_{sr} \right] \delta(c_{is}, c_{jr}),$$

where $A^s_{ij}$ denotes the adjacency matrix for network layer $s$ (i.e., that associated with time window $\tau_s$), $\gamma$ is again a tunable “resolution parameter,” $k^s_i = \sum_j A^s_{ij}$ is the weighted node degree for node $i$ in layer $s$, $2m_s = \sum_{ij} A^s_{ij}$ is twice the total number of edges in layer $s$, $\delta(m,n)$ is again a Dirac delta function, $C_{sr} = \delta(s,r - 1) + \delta(s,r + 1)$ defines the coupling between consecutive layers and $C_{sr} = 1$ if only if $r = s \pm 1$ (otherwise $C_{sr} = 0$), and $\{c_{is}\}$ are the integer indices that indicate the community for each node $i$ in each layer $s$.

As opposed to single-layer modularity, multilayer modularity involves two parameters $\gamma$ and $\omega$. We study modularity-optimizing communities for a range of parameter values $\gamma \in [\gamma_{\text{min}}, \gamma_{\text{max}}]$ and $\omega \in [\omega_{\text{min}}, \omega_{\text{max}}]$ to study the multiscale organization of clusters. This approach efficiently explores clustering phenomena at multiple spatial and temporal scales, identifying at which scales clustering is most prevalent and at which scales it is nonexistent. We use a two-dimensional version of the previously-mentioned CHAMP algorithm [56] to efficiently explore this parameter space.

5.7 Calibrating Multilayer Modularity with Layer-Matching Single-Layer Modularity

In order to calibrate the multi-layer modularity we consider a simple extension to the single-layer modularity optimization algorithm that allows it to apply to multi-layer networks. This is the layer-matching extension. This begins by applying the single-layer algorithm separately to each layer. Then, we do a forward pass through the layers, and any time two communities in adjacent layers share more than a certain fraction $\delta$ of nodes, they are assigned the same layer. In this way, stable communities tend to keep the same ID over layers.
However, the single-layer modularity method does not use the information from adjacent layers when constructing the communities, so when the communities are less obvious it will be less likely to detect them.

6 Acknowledgements

The authors acknowledge motivation for this study as well as valuable discussions with Kerry Bloom, Josh Lawrimore, and Paula Vasquez.

7 Competing Interests

The authors declare no competing interests.

References

[1] Benjamin Albert, Julien Mathon, Ashutosh Shukla, Hicham Saad, Christophe Normand, Isabelle Léger-Silvestre, David Villa, Alain Kamgoue, Julien Mozziconacci, Hua Wong, et al. Systematic characterization of the conformation and dynamics of budding yeast chromosome xii. *J Cell Biol*, 202(2):201–210, 2013.

[2] Elnaz Alipour and John F Marko. Self-organization of domain structures by dna-loop-extruding enzymes. *Nucleic acids research*, 40(22):11202–11212, 2012.

[3] Jean-Michel Arbona, Sébastien Herbert, Emmanuelle Fabre, and Christophe Zimmer. Inferring the physical properties of yeast chromatin through bayesian analysis of whole nucleus simulations. *Genome biology*, 18(1):81, 2017.

[4] Andrew S Belmont. Mitotic chromosome structure and condensation. *Current opinion in cell biology*, 18(6):632–638, 2006.

[5] Andrew S Belmont, Gang Li, Gail Sudlow, and Carmen Robinett. Visualization of large-scale chromatin structure and dynamics using the lac operator/lac repressor reporter system. In *Methods in cell biology*, volume 58, pages 203–222. Elsevier, 1999.

[6] Andrew S Belmont and Aaron F Straight. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends in cell biology*, 8(3):121–124, 1998.

[7] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
[8] Chris A Brackley, Benno Liebchen, Davide Micheleletto, Francois Mouvet, Peter R Cook, and Davide Marenduzzo. Ephemeral protein binding to dna shapes stable nuclear bodies and chromatin domains. *Biophysical journal*, 112(6):1085–1093, 2017.

[9] Irineo Cabreros, Emmanuel Abbe, and Aristotelis Tsirigos. Detecting community structures in hi-c genomic data. In *Information Science and Systems (CISS), 2016 Annual Conference on*, pages 584–589. IEEE, 2016.

[10] Haiming Chen, Jie Chen, Lindsey A Muir, Scott Ronquist, Walter Meixner, Mats Ljungman, Thomas Ried, Stephen Smale, and Indika Rajapakse. Functional organization of the human 4d nucleome. *Proceedings of the National Academy of Sciences*, 112(26):8002–8007, 2015.

[11] Jie Chen, Alfred O Hero III, and Indika Rajapakse. Spectral identification of topological domains. *Bioinformatics*, 32(14):2151–2158, 2016.

[12] Tammy MK Cheng, Sebastian Heeger, Raphaël AG Chaleil, Nik Matthews, Aengus Stewart, Jon Wright, Carmay Lim, Paul A Bates, and Frank Uhlmann. A simple biophysical model emulates budding yeast chromosome condensation. *Elife*, 4:e05565, 2015.

[13] Job Dekker. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *The Journal of Biological Chemistry*, 283(50):34532–34540, 2008.

[14] Job Dekker, Marc A Marti-Renom, and Leonid Mirny. Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. *Nature reviews Genetics*, 14(6):390–403, 2013.

[15] Job Dekker and Leonid Mirny. The 3d genome as moderator of chromosomal communication. *Cell*, 164(6):1110–1121, 2016.

[16] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, 2002.

[17] Zhijun Duan, Mirela Andronescu, Kevin Schutz, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C Anthony Blau, and William S Noble. A three-dimensional model of the yeast genome. *Nature*, 465(7296):363, 2010.

[18] Umut Eser, Devon Chandler-Brown, Ferhat Ay, Aaron F Straight, Zhijun Duan, William S Noble, and Jan M Skotheim. Topologically associating genomic domains in yeast. *Proceedings of the National Academy of Sciences*, 114(15):3061–3070, 2017.

[19] Marina Feric, Nilesh Vaidya, Tyler S Harmon, Diana M Mitrea, Lian Zhu, Tiffany M Richardson, Richard W Kriwacki, Rohit V Pappu, and Clifford P Brangwynne. Coexisting liquid phases underlie nucleolar subcompartments. *Cell*, 165(7):1686–1697, 2016.
[20] JK Fisher, M Ballenger, ET O’Brien, J Haase, R Superfine, and K Bloom. DNA relaxation dynamics as a probe for the intracellular environment. Proceedings of the National Academy of Sciences, 106(23):9250–9255, 2009.

[21] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of chromosomal domains by loop extrusion. Cell Reports, 15(9):2038–2049, 2016.

[22] Geoffrey Fudenberg and Leonid A Mirny. Higher-order chromatin structure: bridging physics and biology. Current Opinion in Genetics & Development, 22(2):115–124, 2012.

[23] Surya K Ghosh and Daniel Jost. How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. PLOS Computational Biology, 14(5):e1006159, 2018.

[24] Anton Goloborodko, Maxim V Imakaev, John F Marko, and Leonid Mirny. Compaction and segregation of sister chromatids via active loop extrusion. Elife, 5, 2016.

[25] Houssam Hajjoul, Julien Mathon, Hubert Ranchon, Isabelle Goiffon, Julien Mozziconacci, Benjamin Albert, Pascal Carrivain, Jean-Marc Victor, Olivier Gadal, Kerstin Bystricky, and Aurélien Bancaud. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. Genome Research, 23(11):1829–1838, 2013.

[26] Caitlin Hult, David Adalsteinsson, Paula A Vasquez, Josh Lawrimore, Maggie Bennett, Alyssa York, Diana Cook, Elaine Yeh, Mark Gregory Forest, and Kerry Bloom. Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus. Nucleic acids research, 45(19):11159–11173, 2017.

[27] Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker, and Leonid A Mirny. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nature Methods, 9(10):999–1003, 2012.

[28] Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950):289–293, 2009.

[29] Sijia Liu, Haiming Chen, Scott Ronquist, Laura Seaman, Nicholas Ceglia, Walter Meixner, Lindsey A Muir, Pin-Yu Chen, Gerald Higgins, Pierre Baldi, et al. Genome architecture leads a bifurcation in cell identity. bioRxiv, page 151555, 2017.

[30] Sijia Liu, Pin-Yu Chen, Alfred Hero, and Indika Rajapakse. Dynamic network analysis of the 4d nucleome. bioRxiv, page 268318, 2018.
[31] K Maeshima, S Hihara, and H Takata. New insight into the mitotic chromosome structure irregular folding of nucleosome fibers without 30-nm chromatin structure. In Cold Spring Harbor symposia on quantitative biology, volume 75, pages 439–444. Cold Spring Harbor Laboratory Press, 2010.

[32] John F Marko and Eric D Siggia. Polymer models of meiotic and mitotic chromosomes. Molecular biology of the cell, 8(11):2217–2231, 1997.

[33] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. science, 328(5980):876–878, 2010.

[34] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R La joie, Leonid A Mirny, and Job Dekker. Organization of the mitotic chromosome. Science, 342(6161):948–953, 2013.

[35] Mark EJ Newman. Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23):8577–8582, 2006.

[36] Chad G Pearson, Paul S Maddox, ED Salmon, and Kerry Bloom. Budding yeast chromosome structure and dynamics during mitosis. The Journal of cell biology, 152(6):1255–1266, 2001.

[37] Benjamin D Pope, Tyrone Ryba, Vishnu Dileep, Feng Yue, Weisheng Wu, Olgert Denas, Daniel L Vera, Yanli Wang, R Scott Hansen, Theresa K Canfield, et al. Topologically associating domains are stable units of replication-timing regulation. Nature, 515(7527):402, 2014.

[38] Indika Rajapakse and Mark Groudine. On emerging nuclear order. The Journal of cell biology, 192(5):711–721, 2011.

[39] Indika Rajapakse, Mark Groudine, and Mehran Mesbahi. Dynamics and control of state-dependent networks for probing genomic organization. Proceedings of the National Academy of Sciences, 2011.

[40] Indika Rajapakse, David Scalzo, Stephen J Tapscott, Steven T Kosak, and Mark Groudine. Networking the nucleus. Molecular systems biology, 6(1):395, 2010.

[41] Scott Ronquist, Geoff Patterson, Lindsey A Muir, Stephen Lindsly, Haiming Chen, Markus Brown, Max S Wicha, Anthony Bloch, Roger Brockett, and Indika Rajapakse. Algorithm for cellular reprogramming. Proceedings of the National Academy of Sciences, 114(45):11832–11837, 2017.

[42] Angelo Rosa and Ralf Everaers. Structure and dynamics of interphase chromosomes. PLOS Computational Biology, 4(8):e1000153, 2008.

[43] Yuji Sakai, Atsushi Mochizuki, Kazuhisa Kinoshita, Tatsuya Hirano, and Masashi Tachikawa. Modeling the functions of condensin in chromosome shaping and segregation. PLoS computational biology, 14(6):e1006152, 2018.
[44] SA Schalbetter, A Goloborodko, G Fudenberg, JM Belton, C Miles, M Yu, Job Dekker, Leonid Mirny, and J Baxter. Structural maintenance of chromosome complexes differentially compact mitotic chromosomes according to genomic context. *Nature cell biology*, 19(9):1071–1080, 2017.

[45] Anthony D Schmitt, Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L Tan, Yun Li, Shin Lin, Yiqing Lin, Cathy L Barr, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell reports*, 17(8):2042–2059, 2016.

[46] Hideki Tanizawa, Osamu Iwasaki, Atsunari Tanaka, Joseph R Capizzi, Priyankara Wickramasinghe, Mihee Lee, Zhiyan Fu, and Ken-ichi Noma. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic acids research*, 38(22):8164–8177, 2010.

[47] Dane Taylor, Rajmonda S Caceres, and Peter J Mucha. Super-resolution community detection for layer-aggregated multilayer networks. *Physical Review X*, 7(3):031056, 2017.

[48] Dane Taylor, Saray Shai, Natalie Stanley, and Peter J Mucha. Enhanced detectability of community structure in multilayer networks through layer aggregation. *Physical review letters*, 116(22):228301, 2016.

[49] Pierre Therizols, Tarn Duong, Bernard Dujon, Christophe Zimmer, and Emmanuelle Fabre. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5):2025–2030, 2010.

[50] Harianto Tjong, Ke Gong, Lin Chen, and Frank Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research*, 22(7):1295–1305, 2012.

[51] Paula A Vasquez and Kerry Bloom. Polymer models of interphase chromosomes. *Nucleus*, 5(5):376–390, 2014.

[52] Paula A Vasquez, Caitlin Hult, David Adalsteinsson, Josh Lawrimore, Mark G Forest, and Kerry Bloom. Entropy gives rise to topologically associating domains. *Nucleic acids research*, 44(12):5540–5549, 2016.

[53] Jolien Suzanne Verdaasdonk, Paula Andrea Vasquez, Raymond Mario Barry, Timothy Barry, Scott Goodwin, M Gregory Forest, and Kerry Bloom. Centromere tethering confines chromosome domains. *Molecular cell*, 52(6):819–831, 2013.

[54] Laura Vian, Aleksandra Pękowska, Suhas SP Rao, Kyong-Rim Kieffer-Kwon, Seolkyoung Jung, Laura Baranello, Su-Chen Huang, Laila El Kathabi, Marei Dose, Nathanael Pruett, et al. The energetics and physiological impact of cohesin extrusion. *Cell*, 173(5):1165–1178, 2018.
[55] Renjie Wang, Julien Mozziconacci, Aurélien Bancaud, and Olivier Gadal. Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. *Current Opinion in Cell Biology*, 34:54–60, 2015.

[56] William H Weir, Scott Emmons, Ryan Gibson, Dane Taylor, and Peter J Mucha. Post-processing partitions to identify domains of modularity optimization. *Algorithms*, 10(3):93, 2017.

[57] Hua Wong, Jean-Michel Arbona, and Christophe Zimmer. How to build a yeast nucleus. *Nucleus*, 4(5):361–366, 2013. PMID: 23974728.

[58] Hua Wong, Hervé Marie-Nelly, Sébastien Herbert, Pascal Carrivain, Hervé Blanc, Romain Koszul, Emmanuelle Fabre, and Christophe Zimmer. A predictive computational model of the dynamic 3d interphase yeast nucleus. *Current biology*, 22(20):1881–1890, 2012.

[59] Zheng Xu, Guosheng Zhang, Fulai Jin, Mengjie Chen, Terrence S Furey, Patrick F Sullivan, Zhaohui Qin, Ming Hu, and Yun Li. A hidden markov random field-based bayesian method for the detection of long-range chromosomal interactions in hi-c data. *Bioinformatics*, 32(5):650–656, 2015.
Figure 1: Full interphase yeast genome using the polymer bead-spring model of [26] (see Section Model) that implements transient SMC-protein-mediated crosslinking of 5 kb domains within the nucleolus on Chromosome XII. The top, middle, and bottom rows depict three values for the binding parameter $\mu$: (top) “very fast” kinetics with very short-lived bonds, $\mu = 0.09$; (center) “intermediate” kinetics but still short-lived bonds, $\mu = 0.19$; (bottom) “slow” kinetics with longer-lived yet still transient bonds, $\mu = 1.6$. (A)–(C) “Snapshots” of 3-dimensional chromosome conformations of all 16 yeast chromosomes during interphase. Blue beads and edges highlight the nucleolus. (D)–(F) Visualization of nucleolar clusters of beads (i.e., 5 kb domains). Domain positions are identical to (A)–(C) and the colors indicate their cluster labels resulting from the network community detection techniques that we present in following sections. (G)–(I) Heatmaps of pairwise distances between beads in the nucleolus provide an analogue of Hi-C bead-bead proximity data (see Methods Section High-Throughput Chromosome Conformation Capture (Hi-C) with Pairwise-Distance Maps). (J)–(L) Depiction of the same pairwise distance information as in (G)–(I), except the enumeration of beads has been permuted to align with communities detected with a naive clustering algorithm (see Section Naive Cluster Detection for Pairwise Distance Maps). Note that the block structure of panel (L) visually suggests the presence of clustering despite there being none (see, e.g., panel F and our results in the following sections).
Figure 2: Gene pairwise interactions are defined as instances in which 5 kb domains are within distance $d^* = 100$ nm of one another. We study the dynamics of gene pairwise interactions with four summary statistics of the 4D simulated dataset: (A) mixing fraction; (B) mean interaction number; (C) mean waiting time; and (D) mean interaction duration (defined in the text), which we plot as a function of the binding parameter $\mu$. Note that the interaction fraction and mean waiting time are both optimized for $\mu \in (0.3, 0.9)$, whereas the mean interaction number and duration decrease monotonically with $\mu$ with apparent sharp transitions near the same intermediate range of $\mu$.

Figure 3: (A) Histograms of pairwise distances between gene domains $\{d_{ij}^{(t)}\}$ at a time $t$ for the three regimes illustrated in Figure 1: fast ($\mu = 0.09$), intermediate ($\mu = 0.19$), and slow kinetics ($\mu = 1.6$). The large peaks near $d_{ij} \approx 50$ nm for $\mu \in \{0.09, 0.19\}$ indicate a large number of domain pairs at this distance and can be interpreted as a “signal” indicating the presence of clustering and identifying which pairs are in the same cluster. The lack of a peak for $\mu = 1.6$ indicates a lack of clustering, despite what is visually suggested by the heatmap in Figure 1(L). (B) Same information as panel (A), except we plot the time-averaged distances across a long time window. Note for $\mu = 0.19$ that the clustering signal (i.e., peak near $d_{ij} = 50$ nm) has disappeared. (C) Same information as panel (A), except we plot the population-averaged gene-gene distances across 10 simulations with random initial conditions. Note the clustering signal cannot be observed for any value of $\mu$. All experiments indicate $d_{ij} \approx 50$ nm as the approximate spatial scale for clustering, and the peak indicates a clustering “signal” can be obscured if one implements population or time averaging.
Figure 4: Distance-threshold-based network construction and connected components. (A–C): Gene-interaction networks with bead positions identical to Figure 1(E) for $\mu = 0.19$. Edges are created between pairs of beads within distance $d^*$ of each other. As $d^*$ increases, more edges are created increasing the network connectivity as measured by the connected components (CC). (D–E): We plot the number of CCs (red curves) versus $d^*$ for $\mu \in \{0.09, 0.19, 1.6\}$. We additionally plot the empirical probability densities for pairwise distances for each $\mu$, taken from Figure 3(A). As $d^*$ increases, the number of CCs monotonically decreases, with sharp drops occurring at values of $d^*$ for which there is a peak in the probability densities. The wide plateau for the CC curve in panel (D) indicates clustering robustly occurs for $\mu = 0.09$ and identifies $d^* = 100$ as an appropriate distance threshold for determining clusters. CC analysis can be used as a network-based cluster-detection algorithm for static (non-time-varying) data; however, it is effective only when the clusters are compact and well-separated from each other. Time-varying data, and/or clusters outside of these “perfect” conditions, demand more advanced techniques.
Figure 5: Multiscale Clustering Analyses with Network Community Detection. (A–C): Networks created from a snapshot of bead positions with $d^* = 130$ for (A) $\mu = 0.09$, (B) $\mu = 0.19$, and (C) $\mu = 1.6$. Colors indicate community labels obtained using the Louvain modularity-optimization algorithm as we describe in Methods Section Multiscale Community Detection for Single-Layer Networks with $d^* = 325$, $\gamma = 10$ and $\omega = 1$. Observe in panel (B) that for these parameters, the community detection algorithm (bead colors) captures the clustering that one perceives more accurately than does the CC analysis. (D–F): Varying the resolution parameter $\gamma$ identifies differently sized communities allowing the identification of multiscale, and possibly hierarchical, clustering in the position data. Meaningful clusters are identified as network communities that are robust to perturbations of $\gamma$. This robustness manifests as plateaus in community properties such as the number of communities (blue curves) and the adjusted mutual information (AMI, red curve) measuring the agreement between the beads' labels according to community detection and CC analysis. Large AMI values indicate a $\gamma$ scale in which the detected communities closely resemble the CCs.
Figure 6: Time-varying Gene Communities, computed using $d^* = 325$, $\gamma = 10$, $\omega = 1$ (A–C): Lifetime of each gene community as a function of the average community size for $\mu \in \{0.09, 0.19, 1.6\}$, respectively. Color indicates relative local density, with red corresponding to high density. (D–F): Cluster plasticity measured by the probability that a randomly selected bead remains in the same community in the next time window, again plotted versus the average number of beads in that community.

Figure 7: We study the dynamics of bead community affiliations through community-level mixing, or cross-communication. We study cross-communication for a large range of $\mu$ by plotting four summary statistics: (A) mixing fraction; (B) mean interaction number; (C) mean waiting time; and (D) mean interaction duration. These statistics are analogous to those studied in Sec. [Gene Pairwise Interaction Is Maximized at a Particular Bond Duration Timescale], but the definitions are changed slightly to address cross communication (see text). These results are qualitatively identical to the results in Figure 2 for pairwise gene interactions, illustrating that cluster formation and inter-cluster exchanges determine the timescale of mixing.