TÉCNICO LISBOA

The impact of burstiness in dynamic networks on disease spreading

Paolo Frazzetto, Elisabeth Griesbauer & Caspar van der Woerd

IST, Network Science, a.y. 2019/2020

December 2, 2019

1 Introduction

Networks are comprised of a set of points (vertices) that have relations to each other (edges), they provide a natural way of modelling a variety of real world phenomena. For example, a network can be used to model routers that are connected to each other or people and the contacts among them. The study of epidemiology focuses on disease spreading, more specifically how a disease spreads (or dies out) over time in a network. The network is typically taken to be a network of people where the disease spreads when they are in contact with each other.

Epidemiology has been a field that has been quite well studied. However, in many studies static networks are considered. This means that the network is flattened over the time, which carries the implicit assumption that interactions are bidirectional and are regularly distributed in time. In practice, this is in general not true, different studies have shown that for an arbitrary pair of people the relation is typically characterized by a high number of interactions in a short-time period followed by a long period of no interactions [1–5]. The bursty nature of interactions among people in time might have significant implications for the study of epidemiology.

In the current report, we discuss the theoretical foundations of temporal networks, disease spreading models and bursty interaction patterns. By means of simulation we explore the impact of bursty interaction patterns in a real-world dynamic network of face-to-face interaction in a museum. The data has been provided by the Sociopatterns project and is discussed in [6]. By comparing the data across days, our goal is to examine how different degrees of burstiness impact disease spreading. More specifically, we want to examine how bursty interactions influence the proportion of infected people in the population, for different transmission rates of the disease.

2 Literature Review

Analysing real-world disease spreading is typically done by considering simple models, the so-called compartment models. These models consider a well-mixed population where all individuals belong to a compartment. All individuals of one compartment have the same state. Within the population everyone can possibly interact with everyone. Most commonly used are the compartments susceptible (**S**), infected (**I**) and recovered (**R**). Using these compartments the SI, SIS and SIR model provide rules for migration of individuals between compartments in case of interaction. In the simplest model, the SI model, individuals move from the susceptible state (**S**) to the infected state (**I**) with an infection rate $0 < \beta \leq 1$. Once infected, individuals will remain infected. The SIS model builds on the SI model by allowing infected individuals to move back to the susceptible state with a rate δ. Similarly, the SIR model allows individuals to move from susceptible to the recovered state (**R**) with a rate δ. Besides looking at the transfer rates β and δ, which are a property of the disease, epidemiologists are interested in the basic reproductive ration R_0. It is defined as $R_0 := \beta\langle k\rangle$, where k is the average number of contacts of an individual (average degree). Intuitively, R_0 gives information about how many individuals one infected person can infect. In case $R_0 < 1$, the disease will die out over time, if $R_0 > 1$ the disease will spread where higher values of R_0 correspond to faster spreading.

A lot of research in disease spreading is done on static networks, which implies that temporal properties of a network are not (fully) considered. Nowadays, the availability of real world temporal data sources increases (e.g. the Sociopatterns project [6]) [7]. The increasing availability and interest in dynamic networks raises the question how disease spreading compares between static and dynamic networks. The temporal properties that are not present in a static network, can have a significant impact on the spreading of a disease. Studies have shown that information is missed when only considering a static network. Several papers dealing with modelling the spreading of epidemics therefore prefer temporal networks, e.g. [1], [7], [4]. According to Rocha and Blondel: "the static approach [...] is more suitable for modeling systems where processes propagate faster than the evolution of the network. Nevertheless, in several networks of interest to epidemiology, the structure may change faster than infectious states, and the static representation becomes deficient" [4]. Contact patterns of a population contain diverse temporal properties that may affect dynamic processes of disease spreading as much as the topology does.

Intuitively a dynamic network is a network in which a time-scale exists, edges and nodes are not present at all times. More formally we can understand a dynamic network as a network where edges are only active in a certain time-frame. Similarly, we could view a dynamic network as a set of static networks for specific times, the (total) static network would then be the flattened network over all these time-specific static networks. Compartment models can be used on dynamic models by assuming that contagion can only occur inside the time window in which the two individuals interact [7].

The potential problems that can occur when the temporal properties are ignored, can easily be understood by considering an example. Consider an undirected temporal network of 3 nodes depicted in figure 1. Where, at time $t = 0$, B and C interact. At $t = 1$ we have that A and C interact. If C is infected this allows the disease to spread from C to A. If A would be infected C could not be infected due to the temporal properties of the network. Now consider a static representation of the same network, obviously the disease could spread from A to B in this case, which is not possible in reality [1]. The lesson here is that static networks can be misleading, as they may provide false information [7].

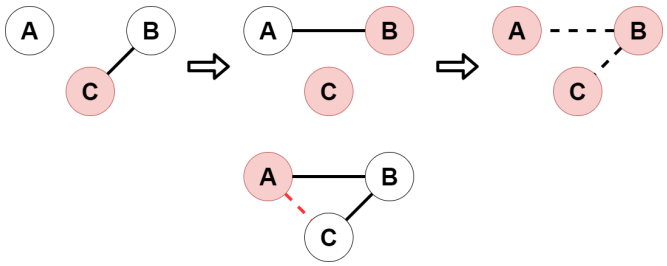


Figure 1: The top shows a dynamic representation of the network, the disease spreads from C to A. If A would be infected, C could not be infected due to the temporal properties of the network. The static representation of the same network depicted at the bottom does allow a disease to spread from A to C [1].

The data that is used for simulation in the current paper has been studied by Isella et al. [6]. In their paper they compare a static representation of the network with the dynamic network. Similarly to the above results they conclude that, although a static network does provide a lot of information, time ordering of events is lost and encoding of causality is not possible.

Another source of potential discrepancies between disease spreading results on static and dynamic networks is the bursty nature of interaction. Several studies have shown that many interactions typically take place in a short time frame, followed by a long period of little to no interactions [2–5]. Initially, researchers assumed that the time between two consecutive events (inter-event time) followed a Poisson distribution. This would imply that the time between two consecutive events is relatively regular [2]. Instead a Pareto distribution of inter-event times is more appropriate. The heavy-tail of this distribution allows for long inter-event times, which would provide a model for the bursty nature as observed in a variety of human activities such as entertainment and work-patterns. A potential explanation for this bursty behaviour is that humans queue tasks based on a perceived task priority. This would imply a Pareto distribution of task waiting time [2].

3 Methods

The goal of this project is to examine the impact of burstiness on disease spreading. To this extent, we study the total proportion of infected nodes, known as the prevalence $\langle i \rangle(\beta)$ where β is a property of the disease and $\langle i \rangle$ is the proportion of infected nodes. We want to answer the question: What is the impact of burstiness in interactions on the prevalence?

Recall the first project, where we extensively described the Sociopatterns data on interactions between museum visitors. The original data consists of 69 unique graphs, corresponding to different days of the experiment. For the current report we use a dynamical representation of the same networks. By means of the SI model, we simulated the spreading of a disease on the network for different values of the infection rate β. The resulting prevalence values were compared across days, by quantifying the amount of burstiness in interactions for each day, we were able to study the impact of burstiness on prevalence. Additionally, the number of infected nodes over time was maintained, to better understand how the prevalence develops throughout the day.

3.1 Disease spreading model

The SI model, which was chosen for its simplicity, was used for simulating the disease spreading on the networks. Since we were dealing with stochastic simulation, we did 100 simulations for a single β value and day, by computing an average over the simulations, we obtained the prevalence. The β value is a property of the disease and was used as a parameter. For each day, 15 values of β were used, the values are equally spaced in the interval $[0.0001, 0.1]$. To guarantee a constant inflow of infected, a prior probability of a node being infected was used, a value of 0.1 was chosen for this.

3.2 Measuring burstiness & sampling of days

The amount of burstiness for a day of the data was quantified using the Fano factor, which is defined as σ^2/μ. Note here, that there is no time dependent factor, instead the variance is the leading term. Therefore, this measure only gives us an indication of how significant the peaks are. However, this measure is not appropriate for assessing the time between peaks, which corresponds to the true meaning of burstiness given the current context. Therefore by means of visual inspection we evaluate the time between peaks in interactions. From the 15 days with the highest Fano factor, 4 days were selected by visually inspecting the amount of time between peaks of interaction. Using the same method, 3 days were selected that were found to have low Fano factor and quite a homogeneous distribution of interactions across the day. The sampling is not only adequate for the burstiness measure, but also for dealing with the computational demands of running many graph simulations.

In the original data the measurements were done on 20 second time frames. The difference between two consecutive time frames was often quite extreme. Using an average over a sliding window of t=10 the temporal data was smoothed. The results and the argument of smoothing are displayed in figure 3 in appendix A. A window size of 10 was chosen to be sufficient for controlling the random error.

3.3 Techniques used & preprocessing

The work was carried out by means of the attached `Python` scripts, in which we designed a class suited for our purposes using dynamic graph library `DyNetx` which is an extension of the `NetworkX` library [8,9].

The original data was preprocessed, such that it was congruent with the input format of `DyNetx`. The raw dataset consists of an edge list for each day - each line has the form (t, i, j), where i and j are the anonymous IDs of the persons in contact, and the interval during which this contact was active is $[t\text{–}20\ sec, t]$. We neglected this real-time reference by setting the first interaction of the day as $t = 0$, then the following interactions occur at discrete time steps of duration 20 *sec* each. Intermediate results were saved by means of `JSON` format text files and `Pickle` files which ensured we did not have to run intensive computations more than once.

4 Results

4.1 Sampled days

In appendix B, the burstiness plots of the sampled days are displayed. Figure 4 displays the number of interactions across time for the 4 days that represent our sample of "high" burstiness. For all these networks, a burstiness score (Fano factor) that is larger than 4 was measured. It can be observed that there are strong peaks where a lot of interactions take place. Figure 5 displays the same graphs for the 3 networks that comprise our sample of "low" burstiness. All burstiness scores are below 1.5, so the Fano factor measures a clear difference between our two groups. The plots also show a relatively homogeneous distribution of interactions compared to the high burstiness group.

4.2 Disease spreading

In Figure 2 the final prevalence of mean infected for $n = 100$ different stochastic realizations is displayed. There is no clear difference observable in prevalence between the networks with low and high burstiness. The prevalence increases quite continuously for an increasing β.

Rather than the final amount of infected, bursts affect the disease spreading behavior over time. The bottom plots of figures 4 and 5 (found in appendix B) show one stochastic realization for every β. By contrasting them with the interaction pattern over time, it is clear that the prevalence rises significantly when a burst in interactions takes place. On the other hand, for days with a more homogeneous interaction pattern, the prevalence increases more constantly.

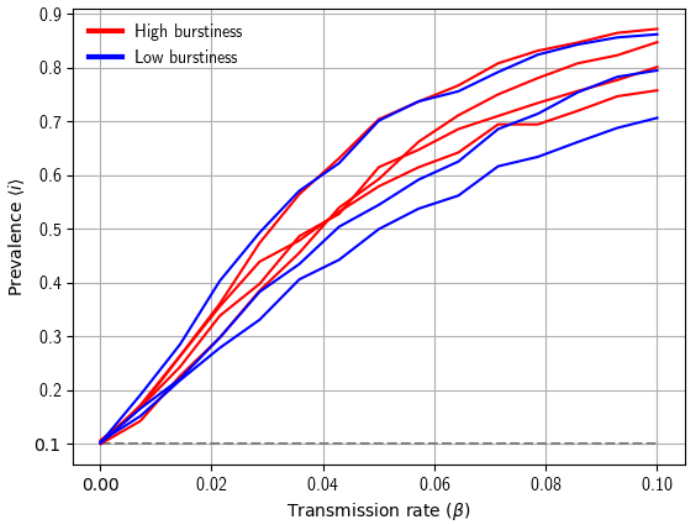


Figure 2: For 15 equally spaced β values, the proportion of infected $\langle i \rangle$ at the final timestamp was computed. For each β the prevalence is the average over 100 simulations. The dotted line represents the minimum expected infected population given that there is a prior probability of 0.1 of being infected for each visitor.

5 Discussion

The relation between rise in prevalence and peaks in interaction during the day can easily be explained. When many interactions take place, there must be many people in the museum. Therefore the burstiness in interactions can also be understood as a heterogeneous presence of people across time. Trivially, a disease can only spread when people are present and interactions take place, therefore a burst in interactions must imply a significant rise in prevalence. Also due to our prior probability of being infected and the use of the SI model (where recovery of disease is not included), the prevalence can only increase over time. Nevertheless, these results do have some implication for prevention policies. The

bursts in interactions are likely to be predictable (e.g. lunch break). Therefore, this burstiness implies that for a small population, as is the case in our data, the critical time where a disease can spread is identifiable. By being aware of this prevention measures can be made time- and space-specific to reduce the chance of spreading during this critical time period. In contrast, for a more homogeneous interaction pattern, the disease gradually spreads, implying that prevention policies should include measures that are effective independent of space and time, such as vaccines.

Interestingly, burstiness in interactions has not been found to lead to a difference in overall prevalence at the end of the day. This implies that a bursty interaction pattern is not a factor that can explain differences in results between dynamic and static networks, given the current model used. Instead, these differences follow from other temporal properties of dynamic networks. It should be emphasized that this conclusion is only valid for the current model where a prior probability of being infected exists and recovery of disease is not included. Also the current networks considered are quite specific, as people typically do not stay in a museum for long, meaning that the people interacting during a peak are likely not present when the next peak takes place. Obviously, many social interactions take place in work/educational environments where this particular property is not true.

For our singular case, the final prevalence is equal for bursty and non-bursty interaction patterns. It would be interesting to examine if this is also the case for different compartment models, different prior probabilities and networks with different properties. If so, the impact of burstiness in disease spreading might be less significant than we had expected.

5.1 Limitations

The current methods rely on a few assumptions that are up to now left implicit. Firstly, the current approach computes runs a single SI-model for each time step. There is no function that determines the probability of infecting someone that takes into account the number of time-frames at which an edge exists. For an arbitrary pair of nodes currently we have $P(A \text{ does not infect } B) = (1-\beta)^n$, where n is the number of time-frames where an edge between A and B exists. If an edge would exist for a long time the probability of a disease transferring along that edge could get quite large. In reality this might not be the most realistic model.

In addition, it should be noted that the current study uses simulations on real-world data which comes with some disadvantages. In the first report we have shown that the networks are somewhat comparable across days. But any of the (potential) results obtained in the current study could not be attributed to the burstiness factor with certainty. Since the networks come from real-world data there might be other factors involved such as disconnected components or other things that are not in our control.

In conclusion, this project has learned us a lot about epidemiology and dynamic networks. Disease spreading on dynamic networks is an interesting topic of research where exciting results may be on the horizon!

References

[1] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS computational biology*, 7(3):e1001109, 2011.

[2] Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207, 2005.

[3] Taro Takaguchi, Naoki Masuda, and Petter Holme. Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PloS one*, 8(7):e68629, 2013.

[4] Luis EC Rocha and Vincent D Blondel. Bursts of vertex activation and epidemics in evolving networks. *PLoS computational biology*, 9(3):e1002974, 2013.

[5] Renaud Lambiotte, Lionel Tabourier, and Jean-Charles Delvenne. Burstiness and spreading on temporal networks. *The European Physical Journal B*, 86(7):320, 2013.

[6] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.

[7] Naoki Masuda and Petter Holme. Predicting and controlling infectious disease epidemics using temporal networks. *F1000prime reports*, 5, 2013.

[8] DyNetx Developers. Dynetx. https://dynetx.readthedocs.io/.

[9] NetworkX Developers. Networkx. https://networkx.github.io/.

Appendices

A Smoothing of burstiness

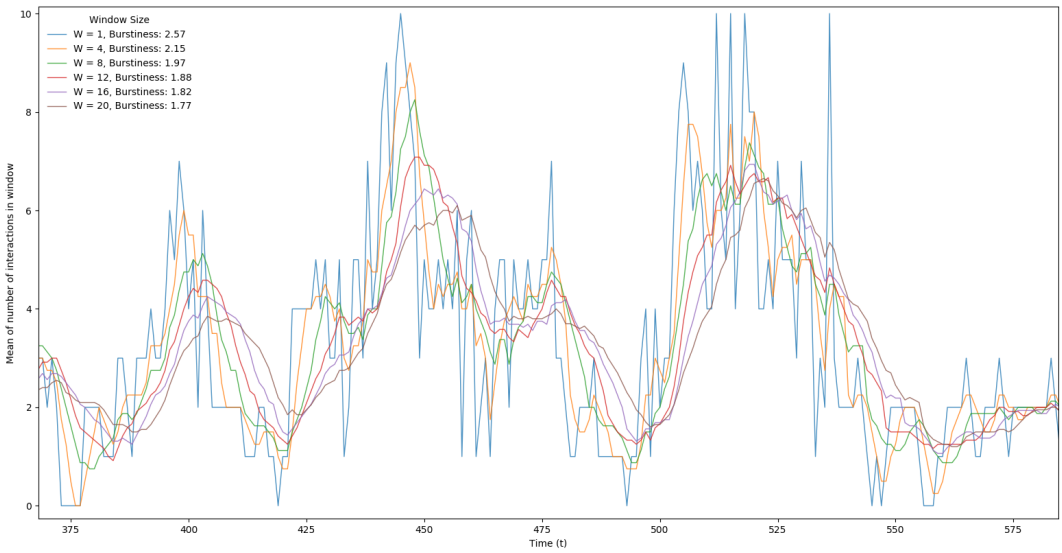
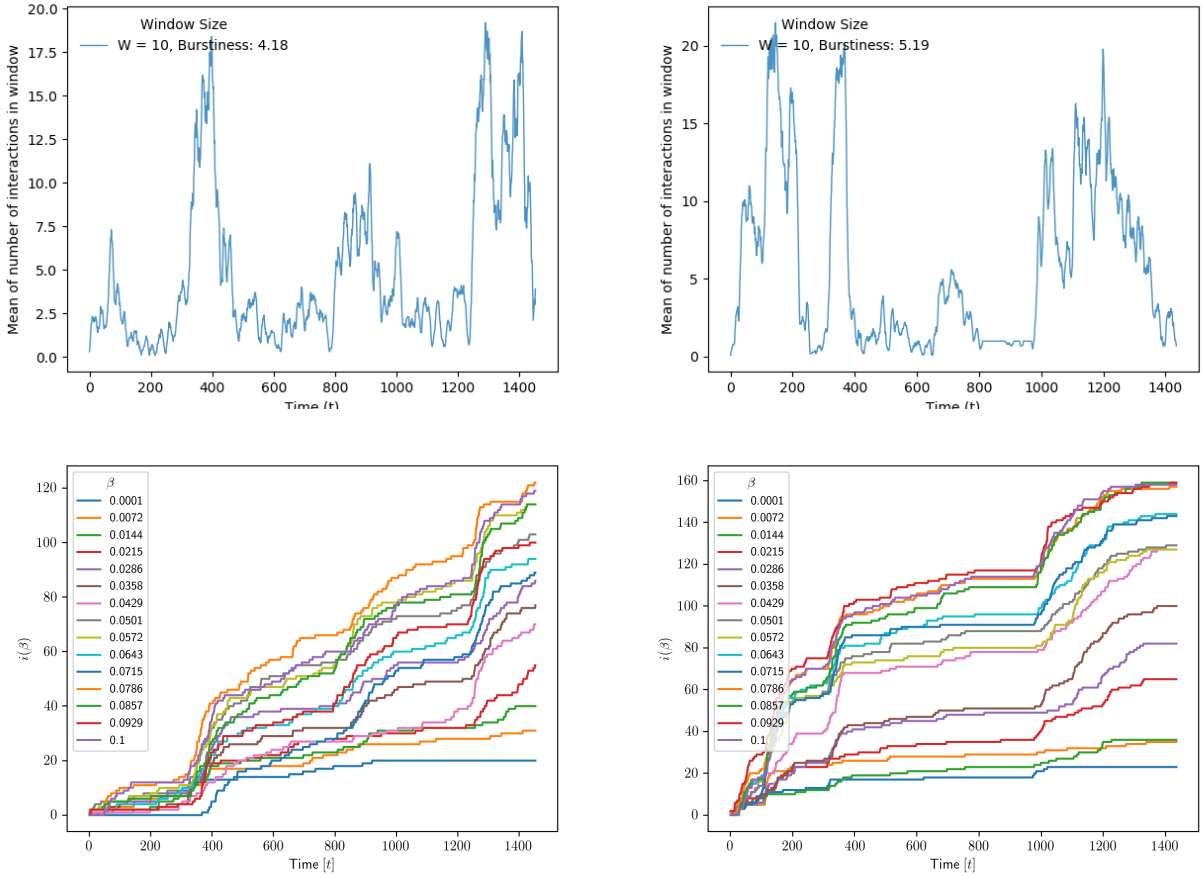


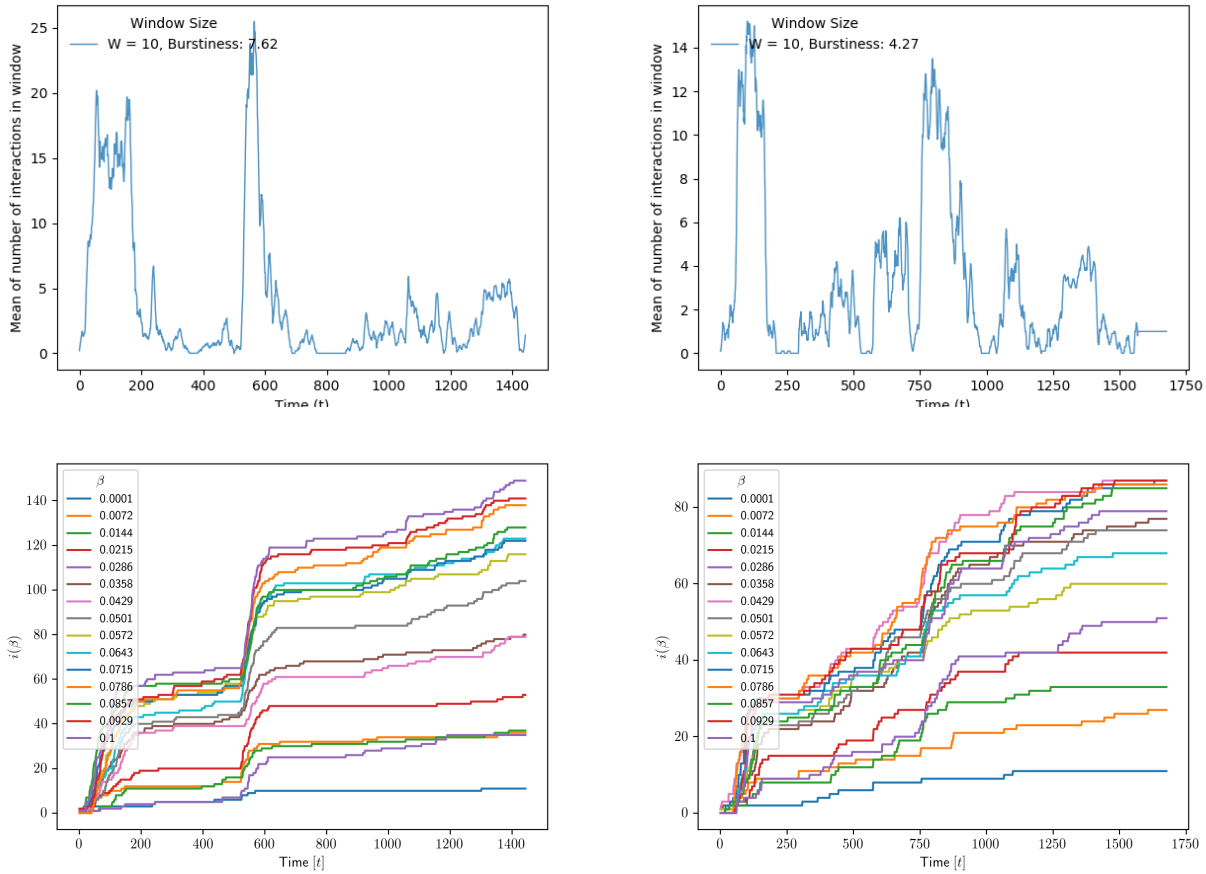
Figure 3: The number of interactions over time was smoothed by computing the mean over a sliding window of size W. For W=1 (no smoothing) there is a lot of random error. A sliding window of W=10 was chosen as this was considered sufficient to deal with the random error

B Sampled graphs

B.1 High burstiness



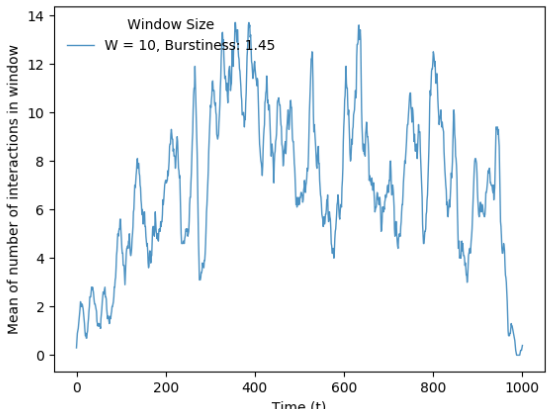
(a) (b)

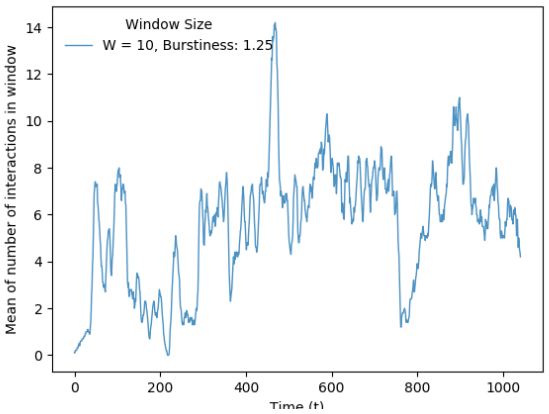


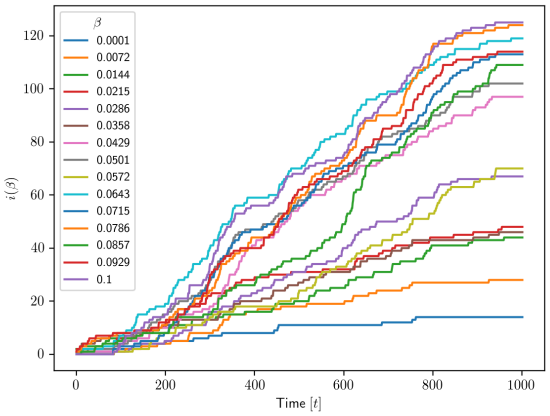
(c) (d)

Figure 4: Each figure represents a day. The sampled days displayed above are considered to have a bursty interaction pattern. The top figure for each day displays the number of interactions (edges) across time (t). The interaction patterns were measured to have high Fano factor (burstiness). By visual inspection these were selected for their short peaks and long times with little to no interactions between these peaks. The bottom plot shows for each day the time evolution for a different stochastic realization with fixed β. Clearly, the prevalence significantly rises when a peak in interactions takes place.

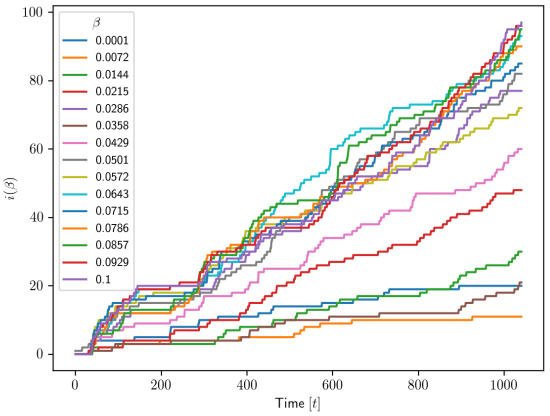
B.2 Low burstiness



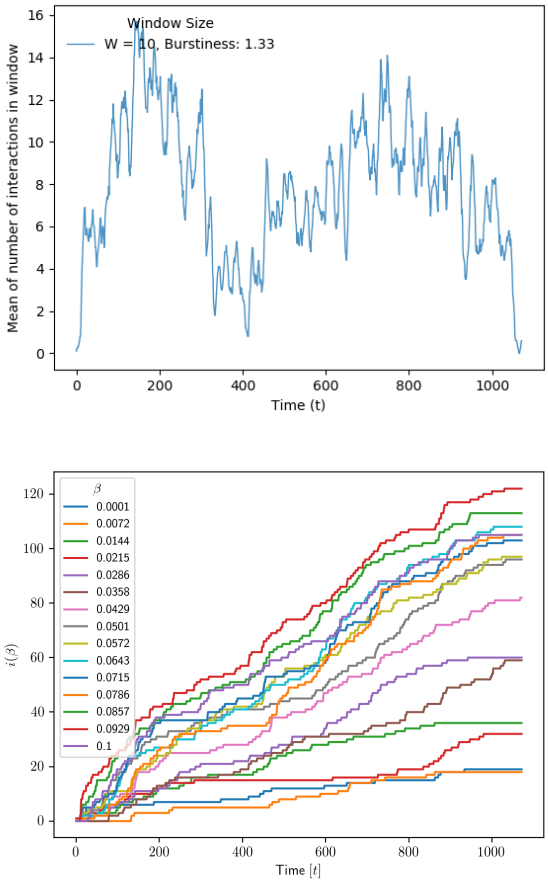




(a)



(b)



(c)

Figure 5: Each figure represents a day. The sampled days displayed above are considered to have a relatively homogeneous interaction pattern. On top: number of interactions (edges) across time (t). The interaction patterns were measured to have low Fano factor (burstiness). By visual inspection these were selected for their relatively constant amount of interactions across time. The bottom plot of each day shows the time evolution for a different stochastic realization with fixed β.