Working papers series

WP ECON 22.12

Uncertainty analysis of contagion processes based on a functional approach

Zonghui Yao
Northeastern University

Dunia López-Pintado
Universidad Pablo de Olavide

Sara López-Pintado
Northeastern University

Keywords: contagion, uncertainty, functional data
JEL Classification: C02, D80.

Department of Economics

Uncertainty analysis of contagion processes based on a functional approach

Zonghui Yao[1], Dunia López-Pintado[2] and Sara López-Pintado[1]

September 2, 2022

Abstract

The spread of a disease (idea or product) in a population is often hard to predict. In reality, we tend to observe only few specific realizations of the contagion process (e.g., the recent COVID-19 pandemic), therefore limited information can be obtained for predicting future similar events. In this work, we use large-scale simulations to study under different exogenous network properties the complete time course of the contagion process focusing on its unpredictability (or uncertainty). We exploit the functional nature of the data, i.e., the number of infected agents as a function of time, and propose a novel non-parametric measure of variance for functional data based on a weighted version of the depth-based central region area. This methodology is applied to the susceptible-infected-susceptible epidemiological model and the small-world networks. We find that the degree of uncertainty of a contagion process is a non-monotonic (increasing/decreasing) function of the contagion rate (the ratio between infectious and recovery probabilities). In particular, maximum uncertainty is attained at the "stable contagion threshold", which represents the parameter conditions for which the endemic/steady state is reaching a plateau as a function of the contagion rate. The effect of the density of the network and the contagion rate are significant and quite similar, whereas the structure of the network, i.e., its amount of clustering/randomness, has a mild effect on the contagion process.

Introduction

The ability to predict the diffusion of a new idea, product, or disease in a population characterized by a complex network of influences and interactions, either virtual or in person, is a fundamental challenge faced by social scientists (e.g., sociologists, economists and epidemiologists) (see Watts, 2004; Vega-Redondo, 2007; Jackson and Yariv, 2007; Ugander et al., 2012; Pastor-Satorras et al.,

[1]Northeastern University, Boston, US. Email: yao.zo@northeastern.edu, s.lopez-pintado@northeastern.edu
[2]Universidad Pablo de Olavide, Sevilla, Spain. E-mail: dlopez@upo.es

2015). Regardless of whether the focus is on the spread of a new "TV series", a "technology" or "biological virus", all of these widely different phenomena have in common that they are hard to anticipate (see Salganik et al., 2006 and the literature cited therein). For instance, in the case of an infectious disease, the rule describing how an agent becomes "infected" by an infectious individual is often perceived as stochastic. Infection depends on numerous factors, such as the time and type of agent's exposure or the status of her immune system, and, therefore the spreading of the disease is subjected to the accumulation of interdependent and uncertain events which leads to random outcomes. Most of the literature on diffusion in networks has focused on studying the long-run or endemic state of the process, either through simulations or by mean-field theory models that approximate the average behavior (Vega-Redondo, 2007). In this paper, we propose a methodology that takes into account the complete time course of contagion and helps discerning features of the model leading to more predictable outcomes.

We use the susceptible-infected-susceptible (SIS) epidemiological model (Pastor-Satorras and Vespignani, 2001) applied to small-world networks (Watts and Strogatz, 1998), which have properties (high clustering and small average path length) that are quite common in the real world (see, for example, Goyal et al., 2006). Our approach is based on a large-scale simulation study and focuses on the functional nature of the data. The unit of study is a random curve ("infected proportion curve") defined as the fraction of infected agents in the population as a function of time. Multiple realizations of this random process generate a sample of infected proportion curves with characteristics that depend on the network structure and the diffusion model. We apply robust non-parametric statistical methods based on depth notions for functional data to describe and analyze the properties of the contagion process (López-Pintado and Romo, 2009; López-Pintado et al., 2014). The variability and, ultimately, unpredictability of this process is estimated by a weighted average of the depth-based central regions of the generated sample of infected proportion curves.

This functional data-based approach allows us to make rigorous estimations about the uncertainty and dynamics of the stochastic model, which depend on the combined effects of the "contagion rate" (the ratio between infectious and recovery probabilities) and the density of the network. However, the estimations are less influenced by the network structure (or randomness). Our unpredictability measure is an non-monotonous (increasing/decreasing) function of the contagion rate. If this rate is below a threshold value, there is little or no diffusion and consequently the process exhibits very low variance. However, when the contagion rate is significantly above threshold, there is widespread diffusion with little uncertainty as the steady state is also reached in a quite predictable manner. Maximum uncertainty is precisely observed at the contagion rate threshold, where convergence to the steady state takes longer and in a less predictable manner.

Methodological approach

The SIS model and small-world networks

The SIS model is typically used to formalize the diffusion of infections that do not confer any long-lasting immunity and thus, upon recovery, individuals become susceptible again (such as the common cold and influenza). It can also be conceived for describing social phenomena such as diffusion of innovations, cultural fads, or economic conventions that share the logic of contagion (e.g. Glaeser et al., 1996; Rogers, 1962; Anderson and May, 1992; Vega-Redondo, 2007).[1]

Formally, in the SIS model a susceptible agent may become infected with a probability ν when interacting with an infectious agent. Reversely, with a probability δ an infected agent can become susceptible again. For simplicity, we will assume a fixed value of δ and vary ν. The key parameter is considered as $\lambda = \frac{\nu}{\delta}$, denoted as the "contagion rate". We assume that there is an initial seed of 10% infected agents which are chosen randomly from the population. The infected proportion of agents in the population at a certain time is denoted as $x(t)$. Given the recurrent transition from susceptible to infected and vice-versa the identities of infected agents vary over time. Also, $x(t)$ is not necessarily monotonous and its particular shape might depend on the parameters of the contagion process and on the realization.

The SIS model is applied to small-world networks which are generated following an approach based on the seminal paper by Watts and Strogatz (1998). To generate the networks we create a ring over S nodes in which each node is connected with its k nearest neighbors (or k - 1 neighbors if k is odd). Each existing link is randomly rewired with a probability r_p, which tunes the nature of the network between that of a unidimensional lattice if $r_p = 0$ and that of a random network if $r_p = 1$ (Supplementary Figure S1). For small, but positive, values of r_p we obtain networks satisfying the small-world properties, i.e., high clustering and short average path lengths. A given network will be characterized by its average density (i.e., k) and randomness (i.e., r_p).

Functional data-based analysis of the contagion process

Consider a network with density k and randomness r_p (Figure 1). Let an initial seed start spreading the disease in the population with contagion rate λ. This determines a Markov process in which the state of the system at a given time is the profile of nodes that are infected versus those that are susceptible. Due to its stochastic nature, multiple realizations of the process generate different infected proportion curves. Figure 1 shows a sample of infected proportion curves for independent draws of the contagion process taking place in the indicated network.

[1] For the diffusion of innovations example it is assumed that the product is adopted with a certain independent probability from the interaction with each neighboring adopter. Also, agents, having adopted, can become potential adopters again if the product breaks, gets lost or becomes obsolete.

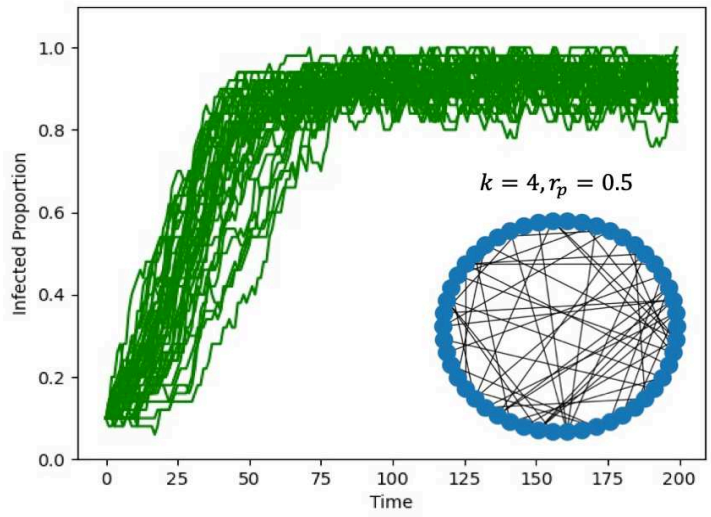

Figure 1: A sample of infected proportion curves. The graph illustrates 40 infected proportion curves simulated given a small world network and the SIS model with $\lambda = 1$ and a 10% random initial seed. The corresponding network (inset) is formed by $S = 50$ nodes, density k=4 and randomness $r_p = 0.5$.

Steady state point and value

Given a contagion process determined by (k, r_p) and λ, let X be a random function, where $X(t)$ is defined as the infected proportion of individuals at time t, with $t \in [0, T]$, and T is the maximum time considered. Let $x_1, x_2, \ldots, x_n$ be a sample of n independent realizations from random function X, i.e., a sample of infected proportion curves. We denote as $M(t)$ to the point-wise median curve, that is,

$$M(t) = Median(x_1(t), x_2(t), \ldots, x_n(t)).$$

The "steady state point" (SSP), denoted by t^*, is the time period where $M(t)$ starts converging to a plateau and fluctuates around this point, i.e., it reaches "approximately" a stationary state. The SSP is interpreted as the moment in time that separates the short-run versus long-run of the process and its calculation requires taking the first derivative of the point-wise median curve (numerical difference approximation) and smoothing it by considering the moving average of a certain order (Supplementary Figure S2). The SSP is defined as the first point in time for which such function is close to zero, e.g., reaches a value lower than a predefined small threshold (fixed at 10^{-7} in our simulation study). Figure 2 (left) illustrates a sample of infected proportion curves, the corresponding point wise median and its SSP. The "steady state value" (SSV), denoted by x^*, is computed as the average of the values taken by $M(t)$ after the steady state point is reached (see x^* in Figure 2). In the simulation study the maximum time is set high enough so that t^* is

smaller than T. In particular, $T = 1000$. Formally,

$$SSV \equiv x^* = \frac{\int_{t^*<t<T} \boldsymbol{M}(t) \quad dt}{\int I(t^* < t < T) \quad dt},$$

where I is the indicator function.

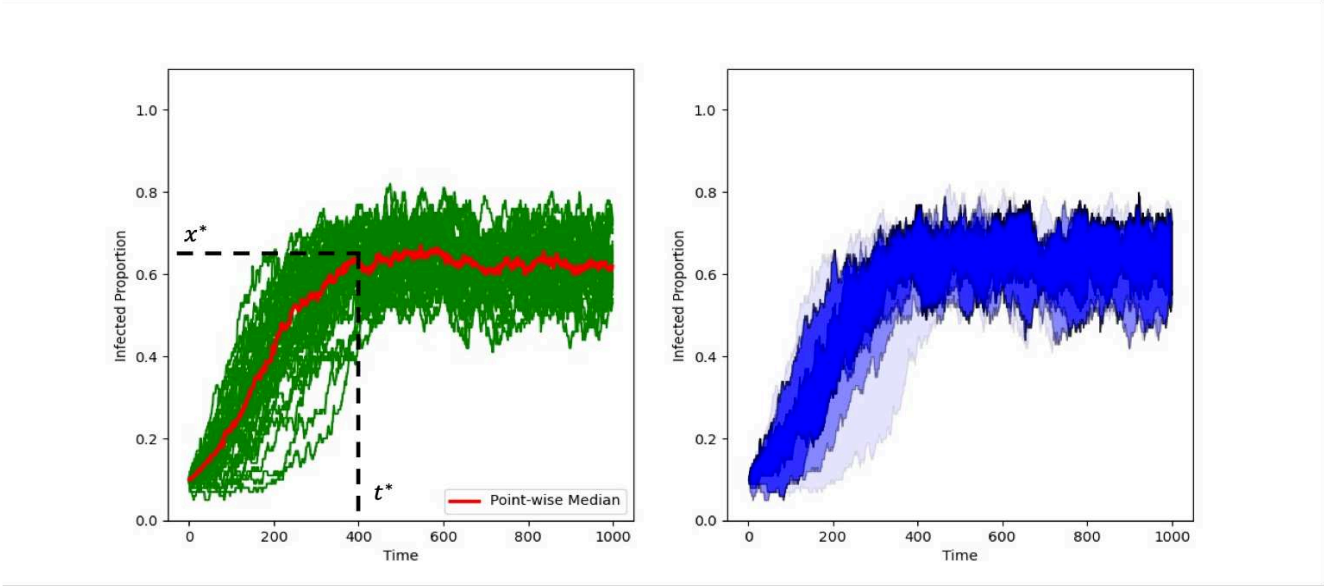

Figure 2: Representation of a sample of 100 infected proportion curves of the contagion model with $\lambda = 1$, k=4 and $r_p = 0.5$. Left. Steady state values: point-wise median (red), SSP=t^* and SSV=x^*. Right. Blue gradient representing the 25%, 50% , 75% and 100% deepest curves in the sample based on the modified band depth (MBD) index.

Depth-based measure of variance

To measure the variability of a sample of curves we apply the notion of functional data depth (see, for example, Fraiman and Muniz, 2001; López-Pintado and Romo, 2009; Sun and Genton, 2011). This approach provides a way of ranking functions in terms of their representativeness with respect to the sample and of defining non-parametric and robust functional statistics. We use the "modified band depth" (MBD) concept, one of the first definitions of depth proposed in the literature, which is based on the regions created by all possible pairs of curves in the sample (López-Pintado and Romo, 2009).

Let $x_1, \ldots, x_n$ be a sample of curves from random curve X defined on the interval $[0, T]$. The (sample) MBD of a function x with respect to the sample $x_1, \ldots, x_n$, is computed as follows:

$$MBD(x) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \Lambda\{B(x; x_{i_1}, x_{i_2})\}$$

where the band is defined as

$$B(x; x_{i_1}, x_{i_2}) = \{t \in [0,T] : \min_{r=i_1,i_2} x_r(t) \leq x(t) \leq \max_{r=i_1,i_2} x_r(t)\},$$

and Λ is the Lebesgue measure on the real line normalized by T.

Therefore, $MBD(x)$ measures the proportion of time, in average and over all possible pairs of curves from the sample, that the curve x is inside the band determined by these pairs of curves. It assigns to the curve x a number between zero and one. The higher this number the more representative the curve x is within the sample. This measure generates a ranking of curves from the deepest (more central) to the least deepest (outlier) in a justified way (López-Pintado and Romo, 2009).[2] Based on the MBD value one can define robust location estimators, such as the median or trimmed mean functions (e.g., the 0.50-trimmed mean is the average of the 50% deepest curves from the sample).

The MBD ranking is also appropriate for measuring the variance of a sample of curves. In particular, the "average p-central region" (ACR_p) determines the variance of a sample of curves as the region/area encompassed exclusively by the proportion p of deepest curves from the sample. The advantage of this concept of variance is that it is intuitive, easy to calculate, and robust, as it neglects the outliers in the sample. Formally, ACR_p based on MBD is defined as:

$$ACR_p = \frac{\int_0^T (\max_{j=1,\ldots,\lceil n\cdot p\rceil} x_{[j]}(t) - \min_{j=1,\ldots,\lceil n\cdot p\rceil} x_{[j]}(t)) \quad dt}{T}$$

where $x_{[1]}, \ldots, x_{[n]}$ are the center-outward ranked curves, with $x_{[1]}$ being the deepest (most central or median) curve, $x_{[n]}$ being the most outlying curve, and $\lceil n \cdot p \rceil$ rounding up to the nearest integer. For example, $ACR_{0.5}$ is the area of the central region/band determined by the 50% deepest curves from the sample and it can be seen as an extension of the standard univariate IQR concept to functional data (see López-Pintado and Romo, 2009; Genton et al., 2014).

An issue with the standard $ACR_{0.5}$ concept is that considering the 50% deepest curves (or p=0.5) instead of the 75% or 25% is quite arbitrary. Moreover, by concentrating only on the deepest subset of curves part of the information obtained in the sample and its ranking is lost. We thus propose a novel measure of variance that weights the more central curves more than the less central ones, but that still contemplates all curves in the sample. A natural way to formalize this idea is to consider a "weighted average" of ACR_p for different values of p. In particular, we consider $p = 0.25, 0.5, 0.75, 1$. To calculate the corresponding weights for each ACR_p, we divide the sample of curves in four groups (or quartiles) according to their MBD values or indexes from highest to lowest. Figure 2 (right) represents by a color gradient the curves from the sample belonging to the different quartiles, where the stronger the colour the lower the quartile. For a specific group (e.g. the group with highest depths, $p = 0.25$; darkest blue color) the weight is the sum of the MBD indexes of the corresponding curves divided by the sum of all depth indexes

[2]For simplicity, we have avoided a notation distinguishing between the population and sample depths (MBD versus MBD_n) as we are always considering the later.

http://www.upo.es/econ

in the sample.

Formally, the "weighted average of central region" ($WACR$) is defined as follows:

$$WACR = \sum_{p\in\{0.25,0.5,0.75,1\}} \alpha_p \cdot ACR_p,$$

where

$$\alpha_p = \frac{\sum_{i=\lceil n\cdot(p-0.25)+1\rceil}^{\lceil n\cdot p\rceil} MBD_{[i]}}{\sum_{i=1}^{n} MBD_{[i]}}$$

and $MBD_{[1]} \cdots MBD_{[n]}$, represent the MBD values of the sample curves, from the deepest to least deepest one. In other words, $MBD_{[i]} = MBD(x_{[i]})$ for $i = 1, ..., n$. By definition, these weights, α_p, add up to one. Note that the dispersion of the deepest curves will be overweighted in the final $WACR$ dispersion value.

Simulation Results

The networks considered for the simulations are formed by S=1000 nodes, 10% of which are randomly infected initially. They have an average degree k, which ranges from 4 to 16, a rewiring probability r_p, which ranges from 0 to 1 and a contagion rate λ that takes values from 0 to 2. For every network created with parameters (k, r_p), a contagion rate λ, and an initial seed, we run 100 repetitions of the contagion dynamics and derive a set of infected proportion curves. We summarize and visualize the results in a schematic way by representing the SSP, SSV and WACR measures defined above, under the different parameter specifications considered.

The density of the network and the contagion rate

We focus here on the joint effect on diffusion of the density of the network and the contagion rate (i.e., k and λ). The results are summarized in Figure 3, where we show, through the intensity of the colours, how the SSP, SSV and WACR values (top, middle and bottom rows, respectively) depend on k and λ, for three levels of the rewiring probability such as r_p=0.01, 0.5 and 1 (see columns left, middle and right, respectively). The case r_p=0.01 has particular interest because it corresponds with a network structure satisfying the small world network properties (high clustering and low average path lengths).

http://www.upo.es/econ

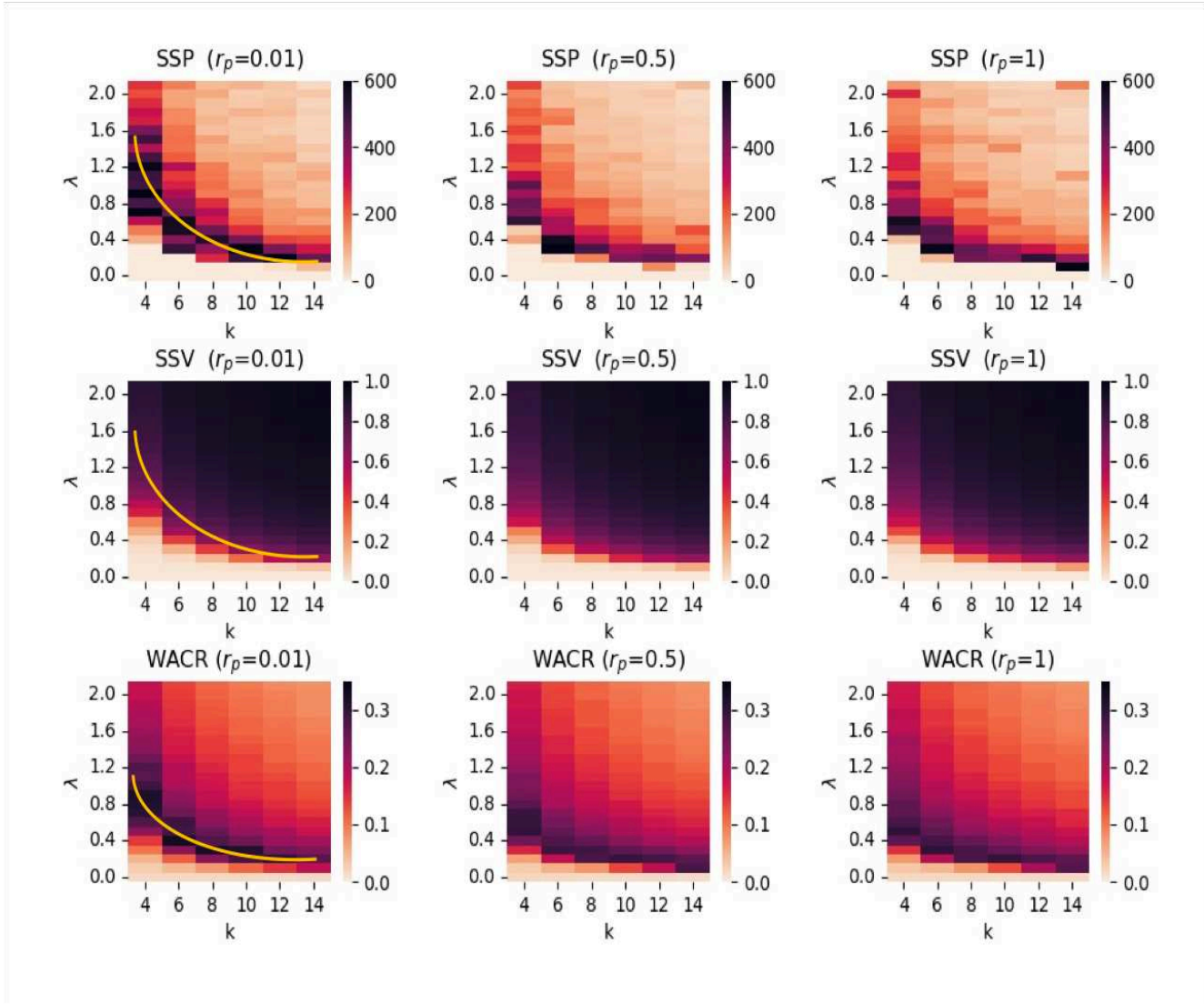

Figure 3: Steady state values and unpredictability of the contagion processes (conditional on k and λ). SSP, SSV and WACR values as a function of λ and k (top, middle and bottom rows, respectively) at three levels of r_p (left, middle and right columns, respectively). The darker the colour the higher the values.

Comparison of the plots (left, middle and right) in every row indicates that the contagion patterns for all network structures is quite similar. In particular, the parameters configurations for which there is little or no diffusion, i.e., SSV is small, (see the "light red" regions in the graphs in the second row of Figure 3) exhibit small convergence time to the steady state (see the same regions in the graphs in the first row of Figure 3) and, also low variance or uncertainty (see the same regions in the graphs in the third row of Figure 3). Similarly, the parameters configurations for which there is a large fraction of infected agents in the steady state (see the "dark red" regions in the graphs in the second row of Figure 3) exhibit also quite fast convergence times (see the same regions in the graphs in the first row of Figure 3) and the variance of the process is relatively low (see the same regions in the graphs in the third row of Figure 3). Finally,

the parameters configuration that are in the intermediate range regarding the endemic state levels (see the "intermediate/mild red" regions in the graphs in the second row of Figure 3) exhibit the longest convergence time to the steady state and the highest variance possible (see the same regions in the graphs in the first and third rows of Figure 3, respectively).

The set of level curves for the SSV as a function of k and λ have a decreasing and convex pattern (i.e., a decreasing hyperbolic shape) which means that there exist some degree of complementarity between the contagion rate and the density of the network regarding the diffusion levels (see one of these level curves in the SSV graph in the second row and first column of Figure 3). This implies that intermediate values of the network density and the contagious rate enhance diffusion (in contrast to more extreme values, as e.g., high value of λ but low k, or vice-versa). Note that the SSP and WACR functions have also the same type of level curves (see Figure 3).

To analyze in further detail the contagion process we focus on the contagion rate as the explanatory variable. In Figure 4 we represent the steady state value as a function of λ for a given network (k=8, r_p=0.01). We define the "stable diffusion threshold", denoted by λ^*, as the inflection point above which the steady state value begins to stabilize, i.e., it changes from a convex/linear function to a concave function. In Figure 4 inset we illustrate a sample of infected proportion curves at the threshold λ^*=0.4. Our simulation results show that it is precisely at this threshold value of λ where the uncertainty of the process is maximized. That is, WACR reaches its maximum value at λ^*, as illustrated in Figure 5 (left column), where the SSP, SSV and WACR are represented as functions of λ for the cases r_p=0.01 and k=4, 8, 12 and 16. We obtain that both the WACR and the SSP are "quasi" bell-shaped (nonmonotonic increasing/decreasing function) with respect to the contagion rate. Also the SSP is maximized at λ^* which implies that the most uncertain process is the one for which the steady state takes longer to reach.[3] In Figure 5 (right column) we also analyze the effect of the density of the network and consider such variable as the explanatory one. The SSP, SSV and WACR are represented as functions of k for the cases r_p=0.01 and λ=0.3, 0.5, 1 and 1.5. We find that WACR and SSP are maximized at the same value of k. When λ takes low values this coincides with the value of k for which SSV starts to stabilize (i.e., the value of k for which SSV, as a function of k, has an inflexion point and changes from convex/linear to concave). For high values of λ the peak of variance is obtained at the lower bound k=4 since SSV, as a function of k, is increasing and concave throughout its the domain. In summary, we observe that the effect on the contagion process of k and λ are quite comparable.

[3]The case k=4 presents a more irregular behavior due to the sparseness of the network. Yet the general findings approximately hold as well.

http://www.upo.es/econ

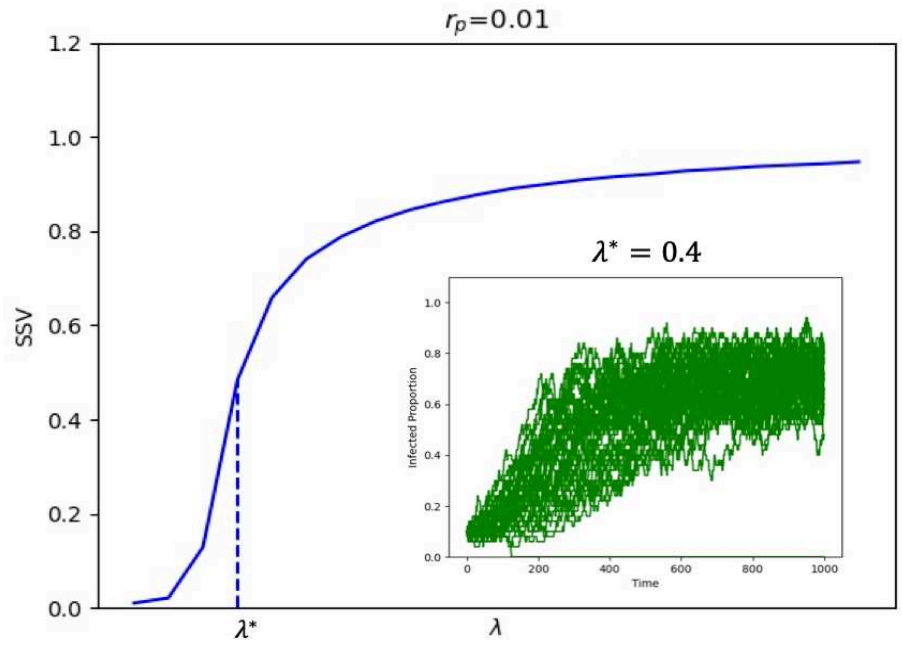

Figure 4: The stable contagion threshold. Representation of SSV as a function of λ in a the case k=8 and r_p=0.01. The stable contagion threshold, λ^*, corresponds with the inflection point of the function. A sample of proportion infected curves at λ^* is shown in the inset of the figure.

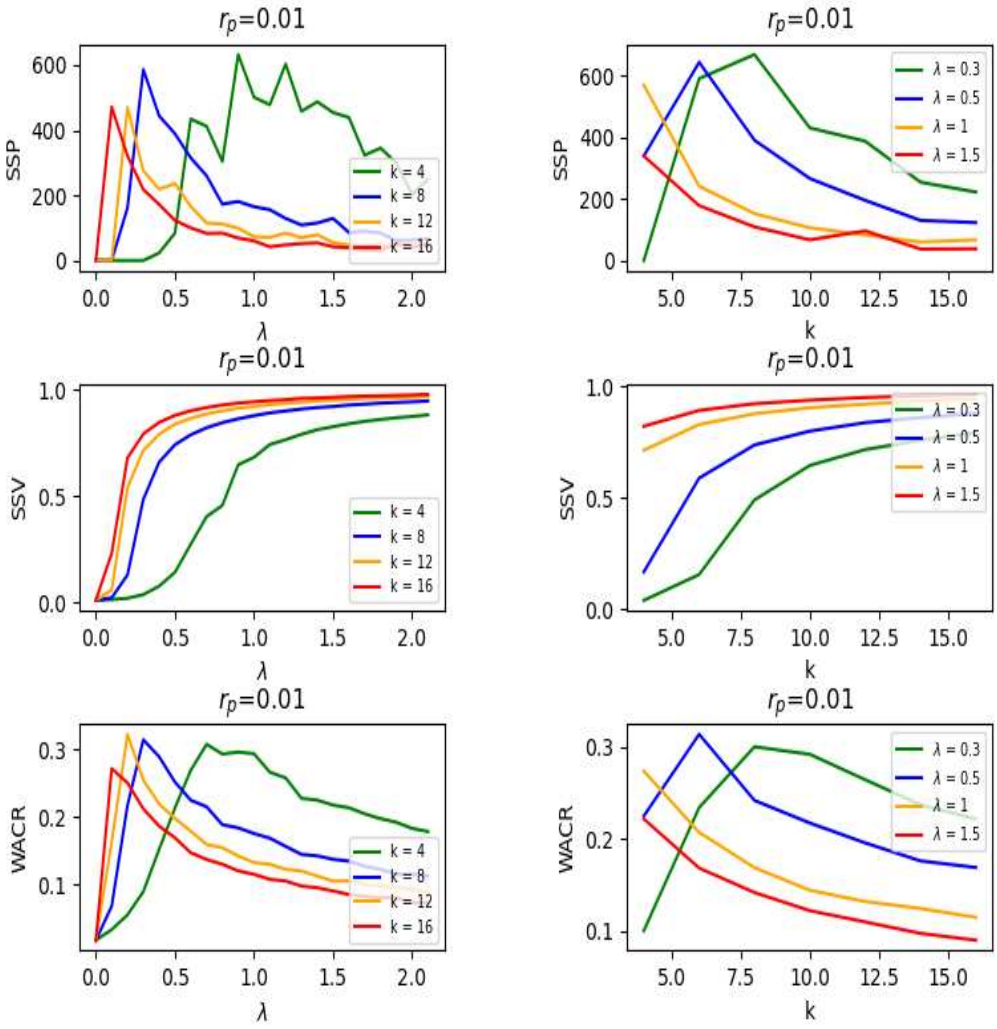

Figure 5: The stable contagion threshold and maximum uncertainty. Left column. Representation of SSP, SSV and WACR as a function of λ for k=4, 8, 12, and 16, and r_p=0.01. Right column. Representation of SSP, SSV and WACR as a function of k for λ=0.3, 0.5, 1, and 1.5, and r_p=0.01.

The randomness of the network and the contagion rate

The results shown in Figure 3 suggest that the effects of the network randomness (or structure) on the contagion process are minor. To confirm this we focus next on the joint effect on diffusion of the rewiring probability (r_p, a key parameter in the small-world networks model) and the contagion rate (λ) and compute the SSP, SSV and WACR as illustrated in Figure 6. We test the process at three different levels of k such as k=4, k=8 and k=16 (left, middle, and right columns, respectively, in the figure). As r_p increases, clustering decreases in the network and so

http://www.upo.es/econ

does the average path length, although the average degree (or density) of the network remains constant. Moreover, the degree distribution becomes more heterogeneous, converging to a Poisson distribution for the random network case (i.e., r_p=1) when the size of the network is sufficiently large (Erdös and Rényi, 1959). We find that the effects of r_p on the three measures analyzed (i.e., SSP, SSV and WACR) are mild, (especially compared to λ's or k's effect). In particular, r_p has either no effect or a slightly positive effect on the SSV (second row). The positive effect might be a consequence of precisely the degree distribution becoming more heterogeneous as r_p increases, which is known to enhance diffusion in the SIS model for purely random networks (Pastor-Satorras and Vespignani, 2001). Surprisingly, despite the fact that the average path length becomes shorter as r_p increases, the convergence time to the stationary state (Figure 6, left column) exhibits a quite erratic behavior. Also the uncertainty of the process presents a very irregular behavior with respect to r_p which cannot be properly observed with the current representations.

SSP, SSV and WACR values are also represented in Figure 7 as a function r_p for the case k=4 and four selected values of λ. We find that, as indicated above, SSV exhibits an increasing or constant trend with respect to r_p. Regarding WACR and SSP, for low values of λ (i.e., λ=0.3 or 0.5), the trend is increasing, but for high values of λ (i.e., λ=1 or 1.5) the trend is decreasing. These observations further indicate that the structure of the network has a minor effect on the uncertainty properties of the contagion process in the particular case of the SIS model.

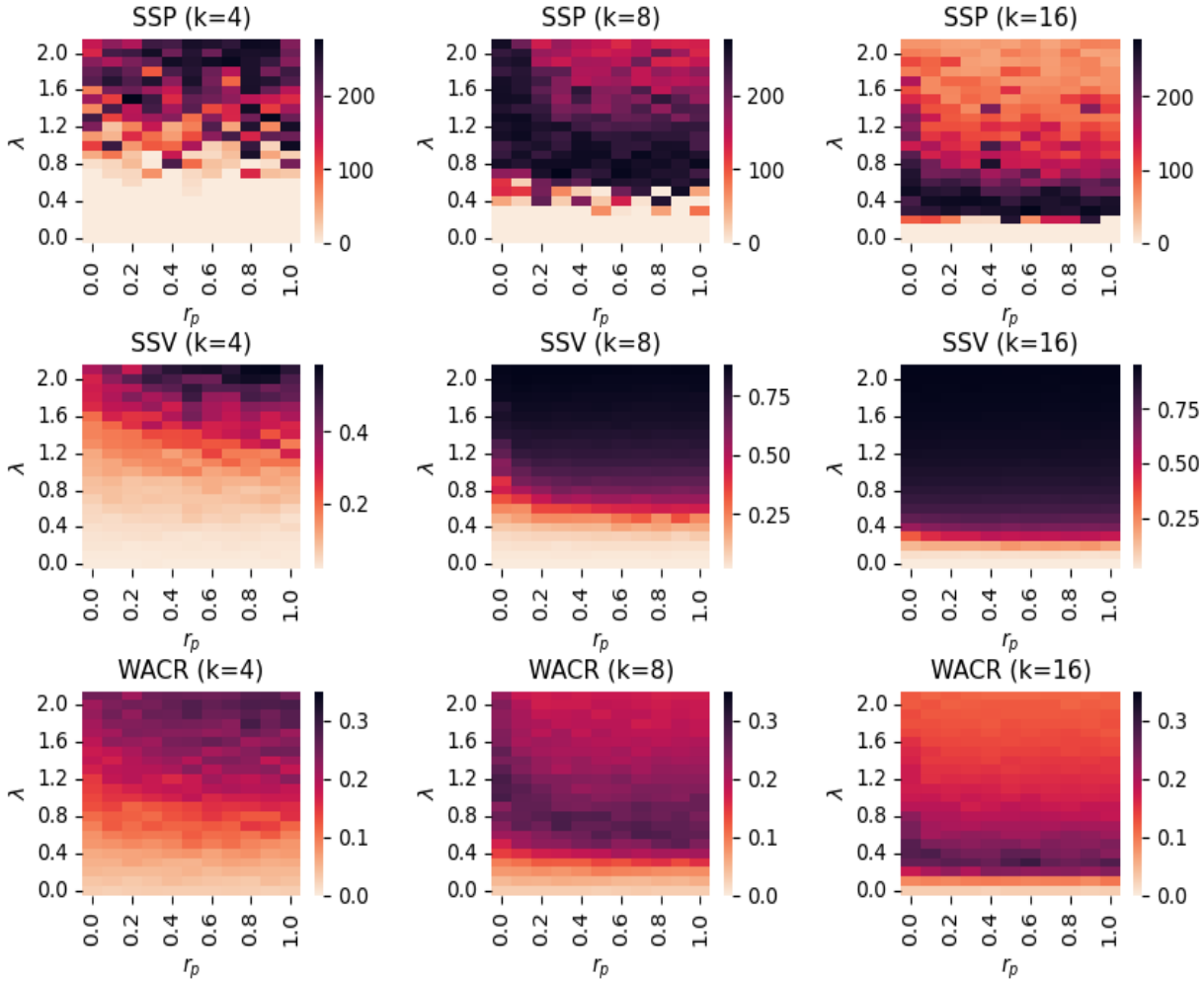

Figure 6: The steady state and unpredictability (conditional on r_p and λ). SSP, SSV and WACR values as a function of λ and r_p (top, middle and bottom rows, respectively) at three levels of k (left, middle and right columns, respectively). The darker the colour the higher the values.

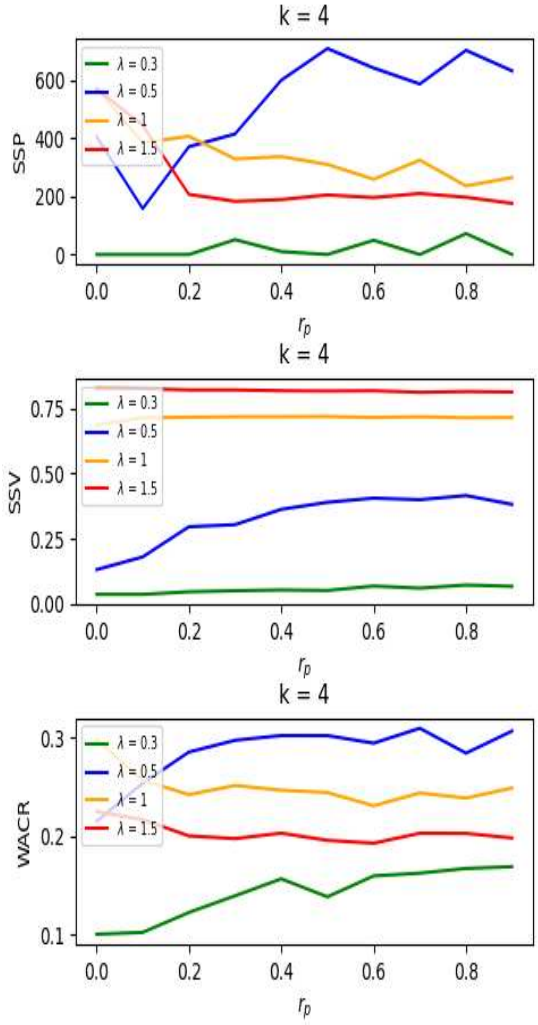

Figure 7: Uncertainty and network randomness. SSP, SSV and WACR values as a function of r_p given four values of λ and k=4.

Concentrated seed

In addition to the standard random initial seed procedure, we have also explored the situation in which the seed is concentrated in one part of the network, a setting which might be more reasonable for some real-world examples of diffusion (e.g., the appearance of a disease that mutates from an animal to a human virus in a particular location). In the simulations, a random node in the network is chosen to be infected initially as well as its neighbors, neighbors of neighbors, etc, until a 10% fraction of infected nodes is reached. Exploratory investigation on this extension shows that qualitatively the results are quite similar to those obtained in the benchmark case. Nevertheless, for structured or close to structured networks (i.e., r_p close

to zero) and for a wide range of parameter specifications, the diffusion takes place faster and reaches a larger fraction of the population when the seed is random (Supplementary Figure S3). A possible explanation for this behavior is that in a structured network there is high clustering and thus redundancies in the ways the disease spreads to neighbors which limits the effectiveness of a concentrated seed, versus a random seed. Unlike the benchmark setting, for the concentrated seed case the impact of the network structure, i.e., r_p, on the contagion process is more relevant.

Discussion

In this paper we have studied the unpredictability of a contagion process using a functional data-based analysis. The classical epidemiological literature has concentrated on determining under which conditions there is positive prevalence of contagion for random networks in the long-run state (Pastor-Satorras and Vespignani, 2001) and more recently, also for small-world networks (Liu et al., 2015). The use of mean-field approximations for the theoretical predictions provides accurate results of the average behavior of the contagion dynamics in the long-run for random networks, but not for structured ones (Moore and Newman, 2000). Pair approximations have been applied to refine mean-field approximations in the context of structured networks (Fieberg and Ellner, 2001; Brännström and Sumpter, 2005). Nervertheless, there are still many challenges to overcome in this respect.

There are two main directions in which our paper contributes to previous literature. First, by considering the whole infection curve we study the process in the short-run and not just in the long-run. We believe that focusing on the complete time course of the contagion process is important since early interventions in a potential epidemics are not only critical for preventing it, but they also might significantly affect the network structure and the contagiousness of the disease (e.g., through confinement policies and the enforcement of masks use). Therefore, the "theoretical" long-run might never be actually reached for the initial set of parameters. Second, unlike standard approaches in the literature, we analyze the properties of multiple realizations of the contagion process and not just the average behavior. This allows us to ask questions about the uncertainty and unpredictability of the process and to analyze how much information is missing when focusing on the "representative" average behavior. The recent COVID-19 experience provides an example where this approach could have been useful. Once a virus starts spreading in a particular area, and assuming that the contagious rate is revealed soon after this, it can be analyzed though a simulation study the likelihood that the same spreading behavior of the disease is expected in other similar/comparable areas by measuring the uncertainty in the given setting. Notwithstanding, a primary objective for future research would be to accompany our initial simulation results with theoretical foundations that would provide broader and better justified answers to the above mentioned questions.

One of the advantages of the analyses carried out in this paper is that they can easily be applied to other contagion models and networks and not just the SIS and small-word network

http://www.upo.es/econ

case. These include direct extensions of the SIS model such as the SIR (susceptible-infected-removed) case as well as more sophisticated epidemiological models with additional individual states. Moreover, in the SIS model the probability of becoming infected depends on the absolute number of infected neighbors but there are other models of contagion for which the probability of becoming infected depends on the relative number of infected neighbors, an assumption that is more reasonable in contexts of opinion formation and social persuassion. These models were originally studied in the context of global interactions (Granovetter, 1978) and later on in random networks (Watts, 2002) and in small world networks (Liu and Xiao, 2012). With respect to the network structure one can model homophily in networks and analyze its impact on the diffusion process (Golub and Jackson, 2012; Jackson and López-Pintado, 2013). We have focused on a stylized case as a starting point, however, investigating alternative models of contagion would be a natural way of proceeding with this line of research.

Finally, it must also be stressed that the notion of depth for functional data, used in the current paper as a robust non-parametric tool for estimating the uncertainty or unpredictability of the contagion process, can be applied to other relevant settings. For instance, simulations and depth-based classification methods could help to infer unidentified information when the contagion rate of an infectious disease (or of a certain product) is known, but the network structure is not or vice versa. Then, given the observation of one (or a few) realization(s) of the contagion process and a family of possible network structures the most plausible network or contagion rate can be inferred.

Acknowledgements

We thank the participants at Loyola Behavioral Lab Workshop 2021 for insightful discussions. We are most grateful to JLB for his help and continuous encouragement. Usual disclaimers apply. Dunia López-Pintado and Sara López-Pintado dedicate this paper to Oscar Pintado-Sanjuán.

Funding

Sara Lopez-Pintado and Zonghui Yao were partially supported by NSF grant DMS-2113696 and by NIH grant 1R21 MH120534-01. Dunia Lopez-Pintado was supported by Ministerio de Ciencias, Innovación y Universidades (PID2020-118585GB-I00), Junta de Andalucía (P18-RT-2135) and UPO- FEDER (1263503).

References

Anderson, R. M. and May, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.

Brännström, Å. and Sumpter, D. J. (2005). The role of competition and clustering in population dynamics. *Proceedings of the Royal Society B: Biological Sciences*, 272(1576):2065–2072.

Fieberg, J. and Ellner, S. P. (2001). Stochastic matrix models for conservation and management: a comparative review of methods. *Ecology letters*, 4(3):244–266.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *TEST*, 10(2):419–440.

Genton, M. G., Johnson, C., Potter, K., Stenchikov, G., and Sun, Y. (2014). Surface boxplots. *Stat*, 3(1):1–11.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly journal of economics*, 111(2):507–548.

Golub, B. and Jackson, M. O. (2012). How Homophily Affects the Speed of Learning and Best-Response Dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.

Goyal, S., van der Leij, M., and Moraga-González, J. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412.

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

Jackson, M. O. and López-Pintado, D. (2013). Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(1):49–67.

Jackson, M. O. and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. *The American Economic Review*, 97(2):92–98.

Liu, M. and Xiao, Y. (2012). Modeling and analysis of epidemic diffusion within small-world network. *Journal of Applied Mathematics*, 2012.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.

López-Pintado, S., Sun, Y., Lin, J. K., and Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification*, 8(3):321–338.

Moore, C. and Newman, M. E. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979.

Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63:066117.

Rogers, E. (1962). *Diffusion of Innovations*. Free Press of Glencoe.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

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

Vega-Redondo, F. (2007). *Complex Social Networks*. Econometric Society Monographs. Cambridge University Press.

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

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology*, 30(1):243–270.

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

Supplementary material

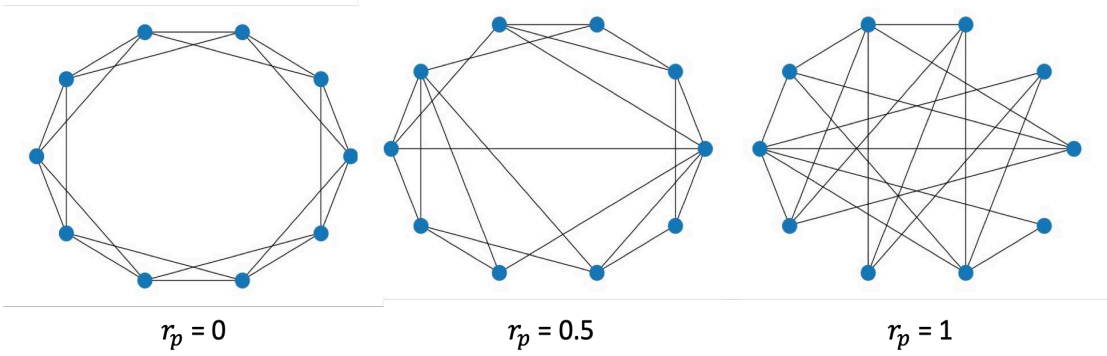

Figure 8: (Supplementary Figure 1) The Watts–Strogatz model of small-world networks where r_p is the rewiring probability.

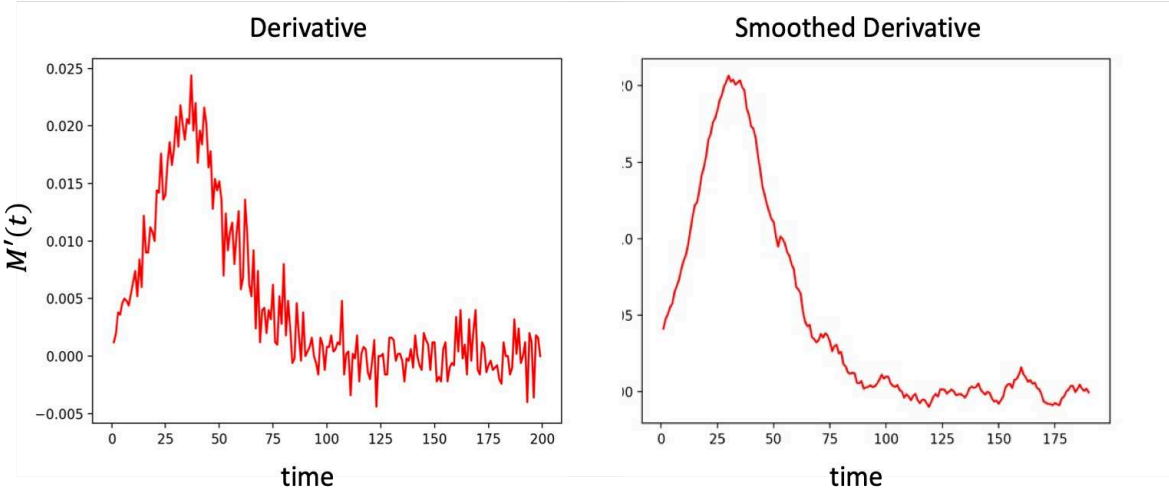

Figure 9: (Supplementary Figure 2) Computation of SSP by numerical difference approximation. Left. Derivative of (point-wise) median curve. Right. Moving average with length 10 sliding window.

http://www.upo.es/econ

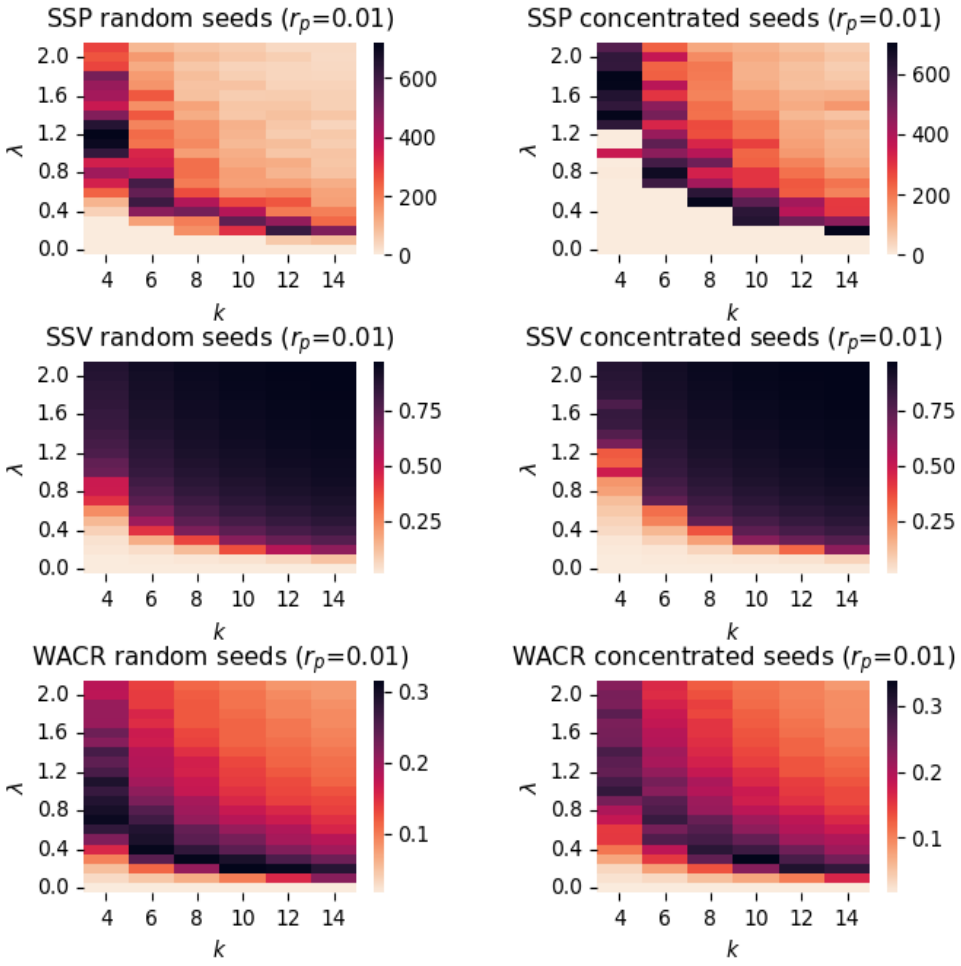

Figure 10: (Supplementary Figure 3) Random versus concentrated seed. SPP, SSV and WACR values as a function of λ and k (top, middle, and bottom rows, respectively) at r_p=0.1 for the random seed versus the concentrated seed cases (left and right columns, respectively). The darker the colour the higher the values.