Testing For Global Covariate Effects in Dynamic Interaction Event Networks

Alexander Kreiss\(^2\), Enno Mammen\(^b\), and Wolfgang Polonik\(^c\)

\(^a\)Institute of Mathematics, Leipzig University, Leipzig, Germany; \(^b\)Institute for Applied Mathematics, Heidelberg University, Heidelberg, Germany; \(^c\)Department of Statistics, University of California at Davis, Davis, CA

ABSTRACT

In statistical network analysis it is common to observe so called interaction data. Such data is characterized by actors forming the vertices and interacting along edges of the network, where edges are randomly formed and dissolved over the observation horizon. In addition, covariates are observed and the goal is to model the impact of the covariates on the interactions. We distinguish two types of covariates: global, system-wide covariates (i.e., covariates taking the same value for all individuals, such as seasonality) and local, dyadic covariates modeling interactions between two individuals in the network. Existing continuous time network models are extended to allow for comparing a completely parametric model and a model that is parametric only in the local covariates but has a global nonparametric time component. This allows, for instance, to test whether global time dynamics can be explained by simple global covariates like weather, seasonality etc. The procedure is applied to a bike-sharing network by using weather and weekdays as global covariates and distances between the bike stations as local covariates.

1. Introduction

One branch in statistical network analysis is concerned with the analysis of so called interaction data. Some references for this type of data are for example Butts (2008), Perry and Wolfe (2013), Matias, RebaKafka, and Villers (2018), and Kreiss, Mammen, and Polonik (2019). Such data typically consists of a dynamic, random network where the vertices are some sort of actors who can interact with each other along the edges of the network. By a dynamic, random network we mean a network which has a fixed set of vertices, that is actors, but the edge set, that is their relations, may change randomly over time. A classical example would be social contact networks: Here the vertices of the network under consideration are people and two persons are connected by an edge if they have the potential to interact with each other (e.g., being in a 50m radius of one another). An interaction, or an event, between two people could be an instance of close contact like the start of a face-to-face conversation. Over the course of the day the potential interaction partners change because people tend to be in different locations during the day (e.g., at work or at home). In addition to the network and the interactions, one typically also observes a set of covariates. Such covariates describe the relation of each pair in the network and can be composed of covariates on actor-level, pair specific covariates or system-wide covariates. The interest lies in modeling the relation between the interactions and the covariates.

We emphasize already here that the edge set does not necessarily need to describe physical (or otherwise established) relations like in the previous example. It is also possible that one imposes a network on the data, for example, by specifying which pairs are relevant for the question of interest. In this case, the edges of the network can be understood as model inclusion dummies, indicating at which point in time which pair is relevant for the study. Such scenarios might be relevant when the interest lies on interactions that are only of interest if other conditions are met, such as the use of social media while using the smart-phone.

We model interaction event data on networks by formulating a counting process model in which the intensity function depends on covariates. In our framework, the covariate process is not required to have short-memory properties. This flexibility allows for models with complex dependence of the covariates on past events. Such models have been studied in both parametric and nonparametric settings, see, for instance, Perry and Wolfe (2013) and Kreiss, Mammen, and Polonik (2019), respectively. In order to illustrate the contribution of this article, we let \(X_{n,ij}(t) \in \mathbb{R}^p\) be local, pair-specific covariate functions. Consider the following model for the intensity \(\lambda_{n,ij}\) of the counting process which counts events from \(i\) to \(j\)

\[
\lambda_{n,ij}(t) := \alpha_0(t) \exp (\beta_0(t)X_{n,ij}(t)).
\]

The most flexible model (fully nonparametric) allows both \(\alpha_0(t)\) and \(\beta_0(t)\) to vary in time. This is studied in Kreiss, Mammen, and Polonik (2019). Perry and Wolfe (2013) assume \(\beta_0 \equiv \text{const.}\) and treat \(\alpha_0\) as nuisance parameter. The case of both \(\beta_0 \equiv \text{const.}\) and \(\alpha_0 \equiv \text{const.}\), which we call a completely parametric model, was used in Kreiss (2021) to study how to test the completely parametric versus the (fully) nonparametric model by using the
$L_2$-distance between a parametric and a nonparametric estimator of the intensity function as a test statistic, similar to what is done in Härdle and Mammen (1993) in nonparametric regression. None of the papers Perry and Wolfe (2013), Kreiss, Mammen, and Polonik (2019), and Kreiss (2021) allows that any entry of $X_{n,ij}(t)$ is the same for all pairs $(i,j)$, thereby rendering the inclusion of a global, system-wide covariate $Z(t)$ impossible. Our work is addressing this issue by including the inclusion of a global, system-wide covariate for tests are relevant because a parametric model allows predicting isane need to use a nonparametric model or whether a parametric model under consideration does not accurately account for this seasonality. Our new test allows to explore whether extending a completely parametric model by allowing for a nonparametric seasonality is producing a meaningful extension of the model. Hence, when discussing the question whether there is a need to use a nonparametric model or whether a parametric model is sufficient, it is natural, to consider intermediate steps between completely parametric and fully nonparametric. We provide the first step for such comparisons. In practice such tests are relevant because a parametric model allows predicting in situations when the global covariates change (like the weather in the next month) while the nonparametric estimate for $\alpha_0$ can only be transferred to other time periods if one assumes that the global covariates remain the same. Lastly, if one has a specific hypothesis about what causes the seasonality, our testing framework provides a methodology to test for this hypothesis.

The complex dependence structure in network settings makes the mathematics behind such types of analysis significantly more challenging. While in standard situations individuals are typically considered to behave independently, such an assumption is rarely plausible in a network set-up. Most of the time it is rather the case that neighboring individuals influence each other and should therefore not be treated as independent. On the other hand it is intuitively clear that this dependence diminishes exponentially as the distance of actors grows in the network: If one is equally influenced by $k$ friends and they, in turn, are equally influenced by their $k$ friends, the impact of the friends of the friends (i.e., of actors of distance 2), is of the order $k^{-2}$. This assumes that no single actor has a disproportionate influence on others (no hubs). Intuitively, if only a single actor influences all things, then observing new actors will not have a major impact. This intuition has been made precise in Kreiss (2021) and we make use of these ideas.

The organization of this article is as follows: In Section 2 we introduce the exact model and present the hypothesis and our suggested test statistic. Afterwards in Section 3 we present the theory for our test statistic. We will discuss the practical implementation in Section 4 and provide a real-world data example as well as a synthetic simulation study. Section 5 concludes. Proofs and other technical details are collected in the supplementary Sections 6–9.

2. Model Specification and Test Strategy

In the following we introduce the exact data generating process used in the following, and formulate the testing problem of interest. Suppose that we observe a sequence of networks $G_{nt} = (V_n, E_{nt})$ for $n \in \mathbb{N}$ over a time interval $t \in [0, T]$, where the deterministic population $V_n$ is of growing size $|V_n| = n$. The edge set $E_{nt} \subseteq \mathcal{V}_n \times \mathcal{V}_n$ is random and time varying. To simplify the notation we identify $V_n$ with $|n| = \{1, \ldots, n\}$. For $i, j \in V_n$, the functions $\alpha_{n,ij}(t) = 1((i, j) \in E_{nt})$ indicate whether $i$ and $j$ are connected at time $t \in [0, T]$ so that $(C_{n,ij})_{ij \in V_n}$ can also be understood as the random, time varying adjacency matrix. Within the population, individuals can interact with each other if they are connected by an edge. For $i, j \in V_n$, we denote by $N_{n,ij}(t)$ the number of interactions between $i$ and $j$ up to and including time $t \in [0, T]$. The processes $N_{n,ij} : [0, T] \to \mathbb{N}$ are thus counting processes. While networks can be directed or undirected, we only consider undirected networks for simplicity. Thus, we assume throughout $C_{n,ij} = C_{n,ji}$ as well as $N_{n,ij} = N_{n,ji}$ for all $i, j \in V_n$. Moreover, also for simplicity, we exclude self-interactions, meaning that we set $C_{n,ii} \equiv 0$ and $N_{n,ii} \equiv 0$ for all $i \in V_n$. We suppose that the network process (via $C_{n,ij}$) and the interactions (via $N_{n,ij}$) are observed. In addition to these we also observe random covariates $X_{n,ij} : [0, T] \to \mathbb{R}^p$ which are specific for the pair $i, j \in V_n$. We are interested in modeling the interactions. We do not model the network process $C_{n,ij}$ and also not the covariate processes $X_{n,ij}$. As outlined in Kreiss, Mammen, and Polonik (2019), one could use our framework for the following more specific model where one observes two interaction processes $N_{n,ij}^-$ and $N_{n,ij}^+$ that define one network process: $N_{n,ij}^+$ jumps if an edge between $i$ and $j$ is added and $N_{n,ij}^-$ jumps if an edge between $i$ and $j$ is removed. Here one would define the network processes belonging to one of the two interaction processes as $C_{n,ij}$ or $1 - C_{n,ij}$ respectively. In this article we will not pursue this setting.

Note lastly that one can also use $C_{n,ij}$ as filters for the researcher to select pairs of interest. While there could potentially be interactions between all vertices, there might be reasons to assume that only specific interactions are relevant for the model (1) at a given time $t$. In this case, the researcher has the flexibility to achieve this by setting all the corresponding $C_{n,ij}(t)$ equal to 1. As was discussed in previous work, the researcher does not have to select the relevant pairs perfectly as long as the selection is not too liberal (see p. 2769 in Kreiss, Mammen, and Polonik 2019).

Throughout we will assume that the array $(N_{n,ij})_{ij \in V_n}$ forms a multivariate counting process with respect to a filtration $(F^p_t)_{t \in [0, T]}$. Unless specified otherwise all counting processes and martingales are understood to be defined with respect to this filtration. Note that by definition of a multivariate counting process no two counting processes jump at the same time (with probability one). As discussed above, the covariates and the interactions are connected through the intensity functions $\lambda_{n,ij} : [0, T] \to (0, \infty)$ for which we assume the proportional hazards
model (see Andersen and Gill 1982; Cox 1972; Andersen et al. 1993; Martinussen and Scheike 2006), that is, we suppose that the intensity function with respect to the filtration \((F_t^n)_{t \in [0,T]}\) is given by

\[
\lambda_{nij}(t) = C_{nij}(t)\alpha_0(t)\Psi(X_{nij}(t); \beta_0),
\]

where \(X_{nij}(t) \in \mathbb{R}^p\) are random covariates depending on time, \(\beta_0 \in \mathbb{R}^q\) is an unknown parameter and \(\alpha_0 : [0, T] \rightarrow [0, \infty)\) is an unknown, deterministic baseline intensity. The link function \(\Psi\) is supposed to be known to the researcher, for example, in a Cox-type model \(p = q\) and \(\Psi(X_{nij}(t); \beta_0) = \exp(\beta_0 X_{nij}(t))\) (see Scheike and Martinussen 2004). In order to assure identifiability, \(\Psi\) or \(X_{nij}\) may not include an intercept, and we have to impose, for example, \(\Psi(0; \beta_0) = 1\). While \(\Psi(X_{nij}(t); \beta_0)\) describes the pair-specific part of the intensity, \(\alpha_0(t)\) can be interpreted as global component of the intensity which applies to all pairs in the system. Our interest then lies in testing whether the baseline \(\alpha_0\) can be adequately modeled by deterministic system wide covariates, that is, covariates that are the same for all individuals, such as weather or economic development. We denote these covariates by \(Z : [0, T] \rightarrow \mathbb{R}^d\). By "deterministic" we here mean "measurable with respect to \(\mathcal{F}_t^n\)" where the measurability assumption on \(Z\) is made for simplicity. Without it, the asymptotic analysis would become significantly more complex. Intuitively the assumption is justified if reliable short-time predictions of future developments of \(Z\) exist, for example, weather forecasts. Since the covariates \(Z\) are supposed to be deterministic and to be the same for the entire network (regardless of its size), we assume also that they do not change with \(n\). Our aim is testing the hypothesis

\[H_0 : \alpha_0(t) = \alpha(Z(t); \theta_0) \text{ for some } \theta_0 \in \Theta,\]

where \(\Theta \subseteq \mathbb{R}^d\) is a suitable parameter space and \(\alpha : \mathbb{R}^d \times \Theta \rightarrow [0, \infty)\) is a known link function. To simplify notation, we let \(\alpha(\theta, t) := \alpha(Z(t); \theta)\). The test statistic we use for testing this hypothesis is along the lines of Härdle and Mammen (1993), that is, we compare a parametric and a nonparametric estimator. In order to define those estimators, we need the following definitions

\[
\begin{align*}
N_n(t) &:= \sum_{i,j \in V_n} N_{nij}(t), \\
\Psi_n(t; \beta) &:= \sum_{i,j \in V_n} C_{nij}(t)\Psi(X_{nij}(t); \beta), \\
C_n(t) &:= I\left(\exists i,j \in V_n : X_{nij}(t) = 1\right), \\
\lambda_n(t; \beta) &:= \alpha_0(t)\Psi_n(t; \beta),
\end{align*}
\]

where the above sums over \(i,j \in V_n\) are understood as sum over all undirected pairs \((i,j) \in V_n \times V_n\) with \(i \neq j\). We will use this notation throughout the article. The process \(N_n\) is again a counting process with respect to \((\mathcal{F}_t^n)_{t \in [0,T]}\) because in our model with probability no two individual processes jump at the same time. The intensity function of \(N_n\) is given by \(\lambda_n(t; \beta_0)\). The function \(C_n\) is an indicator that equals 1 at time \(t\) if there is at least one edge present in the network. We denote by \(M_{nij}(t) := N_{nij}(t) - \int_0^t \lambda_{nij}(s)ds\) the martingale associated with the counting process \(N_{nij}\). Then, \(M_n := \sum_{i,j \in V_n} M_{nij}\) is the martingale associated with \(N_n\) (all with respect to \((\mathcal{F}_t^n)_{t \in [0,T]}\))

The estimators used for our test statistic are as follows. Our parametric estimator is the maximum likelihood estimator

\[
\hat{\theta}_n, \hat{\beta}_n := \arg\max_{\hat{\theta}, \hat{\beta}} \sum_{i,j \in V_n} \int_0^T \log \left(\alpha(\hat{\theta}, t)\Psi(X_{nij}(t); \hat{\beta})\right) dN_{nij}(t) - \int_0^T \alpha(\hat{\theta}, t)\Psi_n(t; \hat{\beta})dt.
\]

For the nonparametric estimator, we first use an initial estimator for \(\beta_0\) not depending on any specific form of \(\alpha_0\). For this we use the partial-maximum likelihood estimator (see Cox 1975; Perry and Wolfe 2013).

\[
\tilde{\beta}_n := \arg\max_{\beta} \sum_{i,j \in V_n} \int_0^T \left[\log \Psi(X_{nij}(t); \beta) - \log \Psi_n(t; \beta)\right] dN_{nij}(t).
\]

Our nonparametric estimator is the smoothed Nelson-Aalen estimator (see Nelson 1969; Ramlau-Hansen 1983; Aalen 1978; Andersen et al. 1993)

\[
\hat{\alpha}_n(t, \tilde{\beta}_n) := \int_0^T K_{h,t}(s)\frac{C_n(s)}{\Psi_n(s; \tilde{\beta}_n)} dN_n(s),
\]

where \(K \geq 0\) is a kernel function and \(K_{h,t}(s) := \frac{1}{h} K\left(\frac{s-t}{h}\right)\). Above we use the convention \(0/0 := 0\). Let further

\[
\alpha_{\text{smooth}}(\theta, t) := \int_0^T K_{h,t}(s)\alpha(\theta, s)ds \quad \text{for all } \theta \in \Theta,
\]

be a smoothed versions of the parametric estimator. Härdle and Mammen (1993) argue that, when comparing nonparametric and parametric estimators, the nature of the test might be dominated by the bias of the nonparametric estimator, this can be avoided by smoothing the parametric estimator. Our test statistic for testing is hence

\[
T_n := \int_0^T \left(\hat{\alpha}_n(t, \tilde{\beta}_n) - \alpha_{\text{smooth}}(\hat{\theta}_n, t)\right)^2 w(t)dt,
\]

where \(w : [0, T] \rightarrow [0, \infty)\) is a weight function with \(\tau := \text{supp}(w) \subset (0, T)\), where \(\text{supp}(w)\) denotes the closed support of \(w\). Thus, \(w\) cuts off the boundary and therefore we may ignore possible boundary issues of the kernel type Nelson-Aalen estimator. We consider the boundary cutoff to be the main role of \(w\), thus, we will later in the simulation choose \(w(t) := I(t \in [\delta, T - \delta])\) for some small \(\delta > 0\).

### 3. Main Results

Here we present the main theoretical result of the article, which states the asymptotic behavior of \(T_n\) under the null-hypothesis and under local alternatives. The assumptions needed for this result to hold are presented and discussed in detail in Section 3.2.
3.1. Main Result

In order to state our main result, we introduce the following notation, where by Assumption (VX) below, all these quantities are well defined:

\[ \mu_n(t; \beta) := \mathbb{E} \left( \Psi(X_{n,ij}(t); \beta) \mid C_{n,ij}(t) = 1 \right), \]
\[ p_n(t) := p_n(\pi(t) = \mathbb{P}(C_{n,ij}(t) = 1) \]
\[ N := m p_n \left( \int_0^T \int_0^T K_{h,t}(s) \frac{w(t)}{\pi(s)} ds dt \right)^{-1}, \]

where \( m := \binom{T}{2} \) denotes the total number of undirected pairs and \( p_n \) and \( \pi \) satisfy:

Assumption (SP): (Sparsity) We assume \( p_n > 0, \pi(t) \geq 1 \), and \( \pi \) continuous with \( 0 \leq \pi(t) \leq 1 \) for all \( n \geq 1 \) and \( t \in [0, T] \).

Note that assuming \( \pi(t) \geq 1 \) is no restriction when \( 0 < \int_{t=0}^{T} \pi(t) \), because in that case we may simply rescale \( p_n \).

Observe that the expected number of undirected pairs at time \( t \) equals \( m p_n(\pi(t)) \), and that we explicitly allow that \( p_n \rightarrow 0 \) so that our set-up includes sparse networks in which the expected number of edges \( m p_n(\pi(t)) = O(n) \). Finally, \( N \) can be interpreted as a weighted time average of these expected numbers, see also Remark 3.2. Also note that \( m p_n(\pi(t)) = \mathbb{E} \bar{Y}_n(t, \beta) \).

Theorem 3.1. Suppose that Assumption (SP) and all the assumptions from Section 3.2 hold. Further, assume that model (1) holds with \( \alpha_0(t) = \alpha(\theta_0, t) + c_n \Delta_n(t) \), where \( c_n = \left( N \sqrt{h} \right)^{-1/2} \), and \( \Delta_n \) is uniformly bounded and continuously differentiable with uniformly bounded derivative (that is, uniformly in both \( t \) and \( n \)). Then, as \( n \rightarrow \infty \),

\[
N \bar{Y}_n \left( T_n - \frac{A_n}{Nh} - \int_0^T \left( \int_0^T K_{h,t}(s) c_n \Delta_n(s) ds \right)^2 w(t) dt \right)
\]

\[ \frac{d}{dN} \rightarrow \mathcal{N}(0, 1), \]

where with \( f_n(r, s) := \int_0^T h K_{h,t}(s) K_{h,t}(r) w(t) dt, \gamma := \int_0^T \int_0^T K_{h,t}(s) \frac{w(t)}{\pi(s)} ds dt \), and \( K(2) := \frac{1}{2} \|K \|_2^2 \),

\[ A_n := \frac{1}{N \gamma^2} \sum_{i,j \in V_n} \int_0^T f_n(r, s) \left( \frac{1}{\pi(r) \mu_n(r; \beta_0)} \right)^2 dN_{n,ij}(r), \]
\[ B_n := 4K^2 \int_0^T \left( \frac{w(s) \mu_n(s)}{\gamma \pi(s) \mu_n(s; \beta_0)} \right)^2 ds. \]

Remark 3.2. We assume in Assumption (B) below that \( r \mapsto \mu_n(r, \beta_0) \) is uniformly bounded from below. Furthermore \( \gamma \rightarrow \int_0^T w(t) / \pi(t) dt > 0 \) if \( \pi \) is continuous. Therefore, \( B_n \) is bounded from below and does not influence the rate of convergence. To illustrate the rate of convergence, we consider the case of \( \pi \) constant. Note that in this case \( \pi \equiv 1 \) and \( p_n = p_n(t) \). Therefore, for \( h > 0 \) small enough,

\[ N = m p_n(\pi(t)) \left( \int_0^T w(t) dt \right)^{-1}, \]

and \( N \) is a scaled version of the expected number of edges.
Yu (2015). There, the dependence of the eigenvalues and eigenfunctions on the variance/squared bandwidth of the Gaussian kernel is explicitly stated. With decreasing bandwidth the mass of the weights moves to later summands, and this leads to an asymptotic normal distribution of these $L^2$ statistics. Furthermore, it shows that the statistic puts more and more weight into later summands which increases the omnibus character of the test. For a comparison of $L^2$-type tests with decreasing bandwidth and with fixed bandwidth in nonparametric regression, see Fan and Li (2000). There, it is shown that these tests cover a class of tests with a large spectrum of properties. This was the reason why we have chosen this class of tests in our mathematical analysis. Furthermore, in Fan and Li (2000) it is also shown that asymptotics with decreasing bandwidth and with fixed bandwidth require different mathematical approaches. In this article we only consider the case of statistics with bandwidths converging to zero. In our proof, we will also not explicitly use the representation of $U$-statistics given above, but we will use martingale central limit theorems instead. Nevertheless, the discussion above is important for the basic understanding of the types of test considered in our work.

### 3.2. Assumptions

In this section we collect the assumptions that we use to prove our main result. We begin with fairly standard assumptions on the data.

**Assumption (VX): (Vertex Exchangeability)**

The tupels $(C_{n,i}, X_{n,i}, N_{n,i})$ are vertex exchangeable, that is, their joint distribution does not change when the vertex labels are permuted.

Since in observational studies names are normally uninformativ, we regard this assumption as not very restrictive. (VX) implies that all quantities indexed by $(i,j) \in V_n \times V_n$ are identically distributed and hence that $\mu_n(t; \beta)$ and $\rho_n\pi(t)$ do not depend on $i,j$.

**Assumption (KBW): (Kernel, Bandwidth, Weight)**

The kernel $K$ is bounded, symmetric about 0 and supported on $[-1,1]$. The weight function $w$ is continuous and bounded with $\mathbb{T} := \text{supp} w \subset (0,T)$. The bandwidth $h$ fulfills $\log m \to 0$, $\sqrt{n} \log m \to 0$.

These assumptions in particular imply that $h \to 0$ and $Nh \to \infty$, which, when interpreting $N$ as the effective sample size, are standard assumptions for kernel smoothing. Choosing $h$ of the form $O(N^{-r})$ for $r > 0$ "small," with the standard choice $r = 1/5$, seems reasonable. Also, the quantity $m = \binom{c}{2}$ can of course be replaced by $n^2$ in these assumptions. Using $m$ makes the origins of the stated assumptions more transparent.

In the following, let $\|f(\cdot)\|_\infty := \sup_{x \in D_f} |f(x)|$ denote the sup-norm, where $f : D_f \to \mathcal{X}$ and $\| \cdot \|$ is a norm on $\mathcal{X}$.

**Assumption (C): (Continuity and Boundness of the Model)**

The link functions $\Psi$ and $\alpha$ are bounded, that is $\|\Psi(\cdot; \beta_0)\|_\infty \leq \infty$ and $\|\alpha(\theta_0; \cdot)\|_\infty < \infty$, and fulfill the following Lipschitz property

$$|\Psi(x, \beta) - \Psi(x, \beta_2)| \leq L_\Psi \|\beta_1 - \beta_2\| \quad \text{and} \quad |\alpha(\theta_1, t) - \alpha(\theta_2, t)| \leq L_\alpha(\|\theta_1 - \theta_2\|).$$

The function $L_\alpha$ satisfies $\|L_\alpha\|_2 < \infty$. Moreover, $\mu_n(\cdot; \beta_0)$ is continuous.

Intuitively, Assumption (C) means that the model is not changing too rapidly neither over time ($\text{continuity of } t \to \mu_n(t; \beta_0)$) nor for different parameters (Lipschitz continuity of $\Psi$ and $\alpha$). The latter is plausible, for instance, if we assume continuously differentiable link functions (with respect to the parameters) and bounded covariates.

**Assumption (P) (Parametric Estimation)**

The estimators $\hat{\beta}_n$ and $\tilde{\beta}_n$ are based on data independent of $(C_{n,i}, X_{n,i}, N_{n,i})_{(i,j) \in V_n \times V_n}$ and satisfy $\|\hat{\beta}_n - \beta_0\| = O_p(N^{-1/2})$, $\|\tilde{\beta}_n - \beta_0\| = O_p(N^{-1/2})$ and $\mathbb{E}(\|\tilde{\beta}_n - \beta_0\|^2) = O(N^{-1})$. Moreover, there is a compact set $K(\beta_0)$ such that $\mathbb{P}(\tilde{\beta}_n \in K(\beta_0)) = 1$.

Since $N$ is the effective number of observations, the assumptions on the rates of consistency of the estimators are standard. The independence assumption is made for simplicity. It holds, for instance, if we use data splitting. We conjecture that the independence assumption can be replaced by a stochastic expansion of the estimators. This has been done in other discussions of test statistics based on the comparison of parametric and nonparametric fits, see for example, Assumption (P1) in Härdle and Mammen (1993).

**Assumption (B): (Local Boundedness)**

$$\sup_{n \in \mathbb{N}} \|\mu_n(\cdot; \beta_0)\|_\infty < \infty \quad \text{and} \quad \sup_{n \in \mathbb{N}} \sup_{\beta \in K(\beta_0)} \|\mu_n(\cdot; \beta)\|^{-1}_\infty < \infty.$$

The upper bound holds, for example, if $\Psi$ is bounded, and the lower bound (an upper bound on the inverse) essentially only means that at no point in time the intensity converges to zero (provided that a link exists). Since we condition on the existence of a link in the definition of $\mu_n$ it might even be plausible to assume that $\mu_n$ does not even depend on $n$ as was argued in Kreiss, Mammen, and Polonik (2019).

For the following assumptions we let $B_n(c) := \{\beta : \|\beta - \beta_0\| \leq \frac{c}{\sqrt{n}}\}$ for an arbitrary constant $c > 0$ and, for any $c_1, c_2 > 0$, define the events $A_n(c_1, c_2) :=$

$$\left\{ \sup_{t \in [0,T], \beta \in B_n(c_2)} \sqrt{p_n(t)} \left( \frac{1}{m p_n(t)} \overline{\Psi}_n(t; \beta) \right)^{-1} - \mu_n(t; \beta)^{-1} \right\} \leq c_1 \sqrt{\frac{\log m}{m}}.$$

**Assumption (CQ): (Continuity and Boundedness of the Model)**

The link functions $\Psi$ and $\alpha$ are bounded, that is $\|\Psi(\cdot; \beta_0)\|_\infty \leq \infty$ and $\|\alpha(\theta_0; \cdot)\|_\infty < \infty$, and fulfill the following Lipschitz property

$$|\Psi(x, \beta) - \Psi(x, \beta_2)| \leq L_\Psi \|\beta_1 - \beta_2\| \quad \text{and} \quad |\alpha(\theta_1, t) - \alpha(\theta_2, t)| \leq L_\alpha(\|\theta_1 - \theta_2\|).$$

The function $L_\alpha$ satisfies $\|L_\alpha\|_2 < \infty$. Moreover, $\mu_n(\cdot; \beta_0)$ is continuous.

Recall that $\mu_n(\cdot; \beta) = \mathbb{E}(\Psi(X_{n,ij}(t), \beta) | C_{n,ij}(t) = 1) = \mathbb{E}\left[ \frac{1}{m p_n(t)} \overline{\Psi}_n(t; \beta) \right]$ and $m = \binom{c}{2}$.

**Assumption (IL): (Law of Large Numbers)**

For any $c_1 > 0$ and any $\delta > 0$, there is $c_1 > 0$ such that, for all $n \in \mathbb{N}$
\( \mathbb{P}(\mathcal{A}_n(c_1, c_2)) \geq 1 - \delta \) and \( \mathbb{P}(\tilde{\mathcal{A}}_n(c_1, c_2)) \geq 1 - \delta. \)

(LL) is essentially a law of large numbers for the hazards \( \Psi(X_{n,ij}(t), \beta) \) in the sense that we require that their averages concentrate around their mean. As discussed in the Supplement (Section 6) this assumption holds under certain mixing conditions.

Denote \( \tilde{C}_{n,ij} := \sup_{t \in [0, T]} C_{n,ij}(t). \)

**Assumption (WC): (Weak Correlation)**

Suppose that

\[
\frac{1}{N^4} \sum_{k_a, a \in V_n} \mathbb{P} \left( \prod_{a=1}^4 \tilde{C}_{n, i_a k_a} = 1 \right) = O(1),
\]

\[
\mathbb{P}(\tilde{C}_{n,12} = 1) + \mathbb{P}(\tilde{C}_{n,23} = 1) + \mathbb{P}(\tilde{C}_{n,34} = 1) = O(1), \tag{9}
\]

\[
\mathbb{P}(\tilde{C}_{n,12} = 1, \tilde{C}_{n,23} = 1) + \mathbb{P}(\tilde{C}_{n,23} = 1, \tilde{C}_{n,34} = 1) + \mathbb{P}(\tilde{C}_{n,12} = 1, \tilde{C}_{n,34} = 1) = O(1), \tag{10}
\]

\[
\sup_{i,j \in V_n, s \in [0, T]} \frac{1}{N^4} \sum_{k_a, a \in V_n} \mathbb{P} \left( \prod_{a=1}^4 \tilde{C}_{n, i_a k_a} = 1 \right) = O(1), \tag{11}
\]

\[
\mathbb{P}(\tilde{C}_{n,ij} = 1) = O(1),
\]

\[
\sup_{i,j \in V_n, s \in [0, T]} \frac{1}{N^3} \sum_{k_a, a \in V_n} \mathbb{P} \left( \prod_{a=1}^3 \tilde{C}_{n, i_a k_a} = 1 \right) = O(1), \tag{12}
\]

\[
\mathbb{P}(\tilde{C}_{n,ij} = 1) = O(1).
\]

To motivate this condition, note that the first term in (9) requires that the probability of an edge being present at some time-point in the interval \([0, T]\) is of the same order as the probability being present at a fixed point \(t \in [0, T]\). This appears to be a reasonable assumption if there are active edges, appearing and disappearing continuously, and inactive edges that never appear. This assumption excludes, for instance, the case where every edge appears exactly once at a random time point and exists for only a short time period. Furthermore, a sufficient condition for the third term in (9) to be bounded is that \(\tilde{C}_{n,12}\) and \(\tilde{C}_{n,34}\) are independent. Note that this involves only disjoint pairs \((1, 2)\) and \((3, 4)\). If there is an overlap as for \((1, 2)\) and \((2, 3)\) appearing in the second term, the additional factor of \(n\) in the denominator allows for strong dependence. This type of assumptions have also been used in Kreiss, Mammen, and Polonik (2019) and Kreiss (2021). For the remaining conditions similar interpretations can be found. The conditions on the conditional probabilities require that no single pair is indicative of the behavior of the entire network. Assumption (WC) may be replaced by the weaker but more technical assumptions presented in Section 7.3 in the Supplement.

For the following technical assumption we consider, for each \(t\) and \(n\), a random distance function between pairs \(d^n_{ij} : V_n \times V_n \to [0, \infty)\), and for \(I \subseteq V_n \times V_n\), set \(d^n_{ij}(I, \delta) := \inf_{(k, j) \in I} d^n_{ij}(k, j, k, l)\), with the convention \(d^n_{ij}(ij, \emptyset) := \infty.\)

**Assumption (mDep): (Momentary-m-Dependence)**

There is a number \(M > 0\) such that

\[
\forall n \in \mathbb{N}, \forall t_0 \in [0, T), \forall I \subseteq V_n \times V_n : \text{Given } \mathcal{F}^n_{t_0},
\]

\[
\sigma \left( (N_{n,ij}(t), C_{n,ij}(t), X_{n,ij}(t))_{(i,j) \in I, t \in [t_0, t_0 + 6h]} \right) \text{ is conditionally independent of}
\]

\[
\sigma \left( (N_{n,ij}(t), C_{n,ij}(t), X_{n,ij}(t)) \right) \mathbb{P} \left( d^n_{ij}(ij, I) \geq M \right) = \mathbb{P} \left( d^n_{ij}(ij, I) \right)
\]

\[
, t_0 - 6h \leq s \leq t_0, t_0 \leq t \leq s + 6h, (i, j) \in V_n \times V_n. \right)
\]

This assumption captures the situation of time developing networks where dependency structures may vary over time. The dependence structures are governed by the distance function \(d^n_{ij}\). An example for \(d^n_{ij}\) would be the length of the shortest path in the network between two edges (if existent, and \(\infty\) otherwise). Intuitively, the above assumption means that, conditional on the past, the immediate short future of far apart processes can be treated as if they were independent. This is plausible if information needs some time to travel through the network. See Supplement, Section 7.1 for a more detailed motivation for this assumption (see also Kreiss 2021).

We call a pair \((i, j) \in V_n \times V_n\) a hub, if it has many close, active neighboring pairs \((k, l)\) during a short period of time. The number of close, active neighboring pairs is defined as the number of pairs which are during any interval of length \(6h\) simultaneously closer than \(M\) (the constant from \(mDep\)) to \((i, j)\). In formulas this is the (choice of \([t - 4h, t + 2h]\) is somewhat arbitrary and made for later convenience)

\[
K^i_j := \sup_{t \in [t - 4h, t + 2h]} \sum_{k, \ell \\ k, \ell \in V_n} \mathbb{P}(d^n_{ij}(kl, \emptyset) < M) \times \mathbb{P}( (k, \ell) \in E_{n,r} \text{ for some } r \in [t - 4h, t + 2h]).
\]

For a given \(n_{hub} > 0\), we call \((i, j)\) a hub if \(K^i_j \geq n_{hub}\). Our weak correlation assumption (WC) allows for correlation between overlapping pairs \((i, j)\) and \((j, k)\). If the pair \((i, j)\) is a hub, this means there are many pairs \((j, k)\) in the edge set during some time interval. Our assumptions allow that during this time interval all such pairs \((j, k)\) are correlated with \((i, j)\). The existence of hubs poses therefore challenges when it comes to the behavior of averages.

**Assumption (NH): (No Hubs)**

There is \(n_{hub} > 0\) such that almost surely \(K^i_j \leq n_{hub}\) for all \(i, j \in V_n\).

(NH) is a simplifying assumption. We will prove our results under the weaker Assumption (HSR) which allows for the existence of hubs. (HSR) is given in the Supplement, in Section 7.2.

Our last assumption appears a bit clumsy. But the reader should note that all statements but the last would be trivially true if we had no conditional but regular expectations. Thus, (BM) below excludes only pathologies in which single pairs react strongly to singular events. By \(N_{n,ij}[a, b]\) we mean the number of
jumps of the process $N_{n,ij}$ in the interval $[a, b]$ and we define for any process $X : [0, T] \to \mathbb{R}$

$$
\int |X(t)| d|M_{n,ij}|(t) := \int |X(t)| dN_{n,ij}(t) + \int |X(t)| \lambda_{n,ij}(t) dt.
$$

Assumption (BM): (Bounded Moments)

Let

$$
\sup_{ij, k, l \in V_n} \mathbb{E} \left( (1 + N_{n,ij}[0, T] + N_{n,kl}[0, T])^2 \right) = O(1),
$$

(14)

$$
\sup_{ij, k, l, k_2, l_2 \in V_n, \ \in [0, T]} \mathbb{E} \left( N_{n,kl_1}[s - 2h, s] N_{n,k_2l_2}[s - 2h, s] \right) \to 0,
$$

(15)

$$
\sup_{ij, k, l, k_2, l_2 \in V_n, \ \in [0, T]} \mathbb{E} \left( N_{n,kl_1}[s - 2h, s] \right) \to 0,
$$

(16)

$$
\sup_{i_1, \ldots, i_4 \in V_n \ \in [0, T]} \int_0^T \mathbb{E} \left( (N_{n,i_2j_2}[t - 2h, t] + h)(N_{n,i_4j_4}[t - 2h, t] + h) \right) \to 0,
$$

(17)

$$
\sup_{i_1, \ldots, i_4 \in V_n \ \in [0, T]} \int_0^T \mathbb{E} \left( \prod_{a=1}^4 N_{n,i_aj_a}[0, T] \right) \to 0,
$$

(18)

The last statement requires also that the number of pairs $(a, b)$ which are closer than $M$ to $(i, j)$ is small. In that case the supremum is of finite order and the condition is reasonable. Assumption (BM) can be replaced by the assumptions given in Section 7.3 in the Supplement.

### 3.3. Proof of Theorem 3.1

The general outline of the proof is simple. However, there are many technical details that are very tedious to handle. We provide a comprehensive treatment of all the details in the Supplement, in Section 9, and show here the main steps only. Note that $C_n(t) = 1$ if one of the processes $N_{n,ij}$ jumps at time $t$. We begin by rewriting the integrand in the test statistic in the following way (recall the definitions of $M_n$ and $N_n$ in the two paragraphs before (2))

$$
\tilde{a}_n(t; \tilde{\beta}_n) - \alpha_{\text{smooth}}(\tilde{\theta}_n, t)
$$

$$
= \int_0^T K_{h,t}(s) \frac{1}{\Psi_n(s; \beta_0)} dN_n(s) - \int_0^T K_{h,t}(s) \frac{1}{\Psi_n(s; \beta_n)} dN_n(s)
$$

$$
+ \int_0^T K_{h,t}(s) \frac{1}{\Psi_n(s; \beta_n)} - \frac{1}{\Psi_n(s; \beta_0)} dM_n(s)
$$

$$
+ \int_0^T K_{h,t}(s) \left( \alpha(\tilde{\theta}_n, s) - \alpha(\tilde{\theta}_n, s) + c_n \Delta_n(s) \right) ds
$$

$$
+ \int_0^T K_{h,t}(s) \left( \frac{1}{\Psi_n(s; \beta_n)} - \frac{1}{\Psi_n(s; \beta_0)} \right) dN_n(s) = I_1(t) + I_2(t) + I_3(t) + I_4(t),
$$

where

$$
I_1(t) := \int_0^T K_{h,t}(s) \frac{1}{\Psi_n(s; \beta_0)} dM_n(s),
$$

(20)

$$
I_2(t) := c_n \int_0^T K_{h,t}(s) \Delta_n(s) ds,
$$

(21)

$$
I_3(t) := \int_0^T K_{h,t}(s) \left( \alpha(\tilde{\theta}_n, s) - \alpha(\tilde{\theta}_n, s) + c_n \Delta_n(s) \right) ds,
$$

(22)

$$
I_4(t) := \int_0^T K_{h,t}(s) \frac{1}{\Psi_n(s; \beta_0)} - \frac{1}{\Psi_n(s; \beta_n)} \frac{\Psi_n(s; \tilde{\beta}_n)}{\Psi_n(s; \beta_0)} dN_n(s).
$$

(23)

With this notation, we obtain that

$$
N \sqrt{h} T_n = N \sqrt{h} \int_0^T \left( \tilde{a}_n(t; \tilde{\beta}_n) - \alpha_{\text{smooth}}(\tilde{\theta}_n, t) \right)^2 w(t) dt
$$

$$
= N \sqrt{h} \int_0^T I_1(t)^2 w(t) dt + N \sqrt{h} \int_0^T I_2(t)^2 w(t) dt
$$

$$
+ \sum_{i=3}^4 N \sqrt{h} \int_0^T I_i(t)^2 w(t) dt
$$

$$
+ N \sqrt{h} \sum_{i,j=1, i \neq j}^4 \int_0^T I_i(t) I_j(t) w(t) dt.
$$

Now, Lemmas 9.7 and 9.8 in the supplementary Section 9.4 show that the terms involving $I_1(t)^2$ and $I_4(t)^2$ converge to zero. Furthermore, by using additionally the Cauchy-Schwarz Inequality and Lemmas 9.6 and 9.9-9.11, we see that also the cross-terms in the second row converge to zero. Thus, the asymptotic behavior of $T_n$ is entirely determined by the terms $I_1$ and $I_2$. The term involving $I_2$ equals exactly what we have to subtract in the formulation of Theorem 3.1 (mind the definition of $c_n$)

$$
N \sqrt{h} \int_0^T I_2(t)^2 w(t) dt = \int_0^T \left( \int_0^T K_{h,t}(s) \Delta_n(s) ds \right)^2 w(t) dt.
$$
Connecting all the previous results, we see that we have to show that 
\[ B_n^{-1/2} \left( \frac{N}{\sqrt{n}} \int_0^T I_1(t)^2 w(t) dt - h^{-1/2} A_n \right) \xrightarrow{d} \mathcal{N}(0, 1) \]
in order to complete the proof of Theorem 3.1. This is exactly the content of Proposition 9.2 which will be shown Section 9.3 in the Supplement.

The main tool for the proof of Proposition 9.2 is Rebolledo’s Martingale Central Limit Theorem (see Theorem 9.1 in the Supplement). In Rebolledo’s CLT a sequence of martingales converges to a Gaussian Limit Process. This requires two ingredients: First, the sequence of processes must be a sequence of martingales and, second, the variation process of these martingales must stabilize. In our application the process of interest will be driven by \( M_n \). \( M_n \) is a martingale by definition of our counting process set-up. This is a time-wise property of the process and is as such not affected by the network structure (which is a spatial property). Since the asymptotic normality is a consequence of this time-wise martingale property, we may expect that the asymptotic distribution is the same as for the nonnetwork case. The second ingredient, however, the stabilization of the variance process, is a spatial property. This is naturally affected by the network structure. Thus, network dependence can possibly jeopardize a stabilization of the variance process. We provide here assumptions under which this does not happen. Note that the speed of convergence is not of direct interest as long as the limit is correct. This task is, unfortunately, much more difficult than it might at first appear.

Therefore, we have to make assumptions on the network itself. Most importantly, this is Assumption (VX) which allows to reduce convergences of sums to correct behavior of covariates as we detailed after Assumption (WC). For the network this is sufficient because we may use the notion of \( \widetilde{C}_{n,i,j} \) which allows to disentangle the network from time. This is a strong assumption but we provide in Section 7.3 a set of alternative assumptions which may replace (WC) and does not make use of \( \widetilde{C}_{n,i,j} \). For the covariates we do not have such a simple solution and therefore we require in Assumption (LL) concentration of the covariates.

4. Numerical Results

4.1. Implementation

When it comes to applying the test in a real-world situation, one needs two independent sets of observation, and one also has to specify a model for the baseline function. The first set of data is used to compute the parametric estimates \( (\hat{\theta}_n, \hat{\beta}_n) \) as in (2) as well as the partial likelihood estimate \( \hat{\beta}_n \) as in (3). Then, the smoothed Nelson-Aalen estimator \( \tilde{\alpha}_n(t, \hat{\beta}_n) \) will be computed based on the second set of data according to formula (4). After specifying a weight function, the test statistic \( T_n \) can be computed as in (5). The critical value of the test is determined according to Theorem 3.1. Since we need the distribution of \( T_n \) on the null-hypothesis, we set \( \Delta_n \equiv 0 \) but we have to estimate \( A_n, B_n, \) and \( N \). All of the following estimates are computed based on the second dataset. For estimating \( A_n \) we use its representation from Proposition 9.2. To estimate \( A_n \), we replace \( \mu_n(r; \beta_0) \) by the corresponding average where \( \beta_0 \) is replaced by \( \hat{\beta}_n \). More precisely, \( \mu_n(r; \beta_0) \) is estimated by 
\[ \left( \sum_{ij \in V_n} C_{n,ij}(r) \right)^{-1} \left( \sum_{ij \in V_n} C_{n,ij}(r) \exp \left( X_n(r) \hat{\beta}_n \right) \right). \]
Furthermore, \( mp_n(r) \) (the expected number of edges) is estimated by the observed number of edges at time \( r \). The number \( N \) can be estimated using the same estimates. In order to compute the integral in the definition of \( N \) and in further quantities, we suppose that the network and the covariates remain constant over known time-intervals, for example, they change at every hour but remain then constant for 60 minutes. Note that so far we have used only the partial likelihood estimator which works well on the hypothesis and on the alternative. The variance \( B_n \) can be estimated using the same conventions with the following addition: To find critical values for the test, we may assume that we are on the hypothesis and hence we estimate the true global intensity \( \alpha_0(t) \) by \( \alpha(Z(t); \hat{\beta}_n) \).

For finding the confidence area mentioned in Remark 3.3 we use the nonparametric estimator \( \hat{\alpha}_n(\cdot; \hat{\beta}_n) \).

4.2. Empirical Application

Here we apply the methodology from Section 2 to bike sharing data, details of the implementation are mentioned in Section 4.1. The data is based on 527 bike stations in Washington D.C. and its surrounding. We consider these bike stations to be the vertices of a network. An interaction from bike station \( i \) to bike station \( j \) happens if someone rents a bike at \( i \) and returns it at \( j \). We consider such data from May 13, 2018 (Sunday) to May 26, 2018 (Saturday). The data is publicly available and can be downloaded from https://www.capitalbikeshare.com/system-data. Since many of the connections are rarely (or never) used, we attempt to model only the frequently used connections. To this end we construct a network of active pairs as follows: There is a link from bike station \( i \) to station \( j \) if there were at least ten bike rides in April 2018 from \( i \) to \( j \). So we only consider pairs that were active at least twice a week on average over a period of one month. This convention is somewhat arbitrary and a full analysis would require a sensitivity analysis with respect to this choice. Also note that this is a directed network. This is no problem because the result in Theorem 3.1 holds analogously for directed networks.

Figure 1 shows a kernel density estimate of the times at which bike rides happened over the period from May 13 till May 26. The scale of the x-axis are days, with day 1 and 8 being Sundays and days 7 and 14 being Saturdays. One clearly sees a different pattern for working days and weekends. Moreover it appears that in the second week there were more bike rides than in the first week. One might suspect that this is due to the weather situation: Figure 2 shows that there was more rain in the first week than in the second week.

The weather information is publicly available from https://www.wunderground.com/ and was collected at the Ronald Reagan Airport in Washington D.C. Even though the Washington D.C. area is large, it is plausible...
that this local weather information is a valid indicator for the
weather in the entire region. Therefore, the temperature and the precipitation are system-wide covariates that are identical for all pairs. Let $T(t)$ denote the log of the temperature at time $t$ (in degree centigrade, there were no negative temperatures) and let $P(t)$ denote the precipitation (in centimeters). As a parametric model for $\alpha_0$ we consider the following ($\theta = (\theta_1, \ldots, \theta_{17})$)

$$
\alpha(\theta, t) = \exp \left( \theta_1 + (T(t) (T(t))^2) \left( \begin{array}{c} \theta_2 \\ \theta_3 \end{array} \right) + (P(t) (P(t))^2) \left( \begin{array}{c} \theta_4 \\ \theta_5 \end{array} \right) + \left( \begin{array}{c} \sin \left( \frac{t \pi}{24} \right) \\ \sin \left( \frac{2t \pi}{24} \right) \end{array} \right) \left( \begin{array}{c} \theta_6 \\ \theta_7 \end{array} \right) \right) 
+ \left( \begin{array}{c} \cos \left( \frac{t \pi}{24} \right) \\ \cos \left( \frac{2t \pi}{24} \right) \end{array} \right) \left( \begin{array}{c} \theta_8 \\ \theta_9 \end{array} \right) + W(t) \left( \begin{array}{c} \sin \left( \frac{t \pi}{24} \right) \\ \sin \left( \frac{2t \pi}{24} \right) \end{array} \right) \left( \begin{array}{c} \theta_{10} \\ \theta_{11} \end{array} \right) + \left( \begin{array}{c} \cos \left( \frac{t \pi}{24} \right) \\ \cos \left( \frac{2t \pi}{24} \right) \end{array} \right) \left( \begin{array}{c} \theta_{12} \\ \theta_{13} \end{array} \right) + \left( \begin{array}{c} \cos \left( \frac{3t \pi}{24} \right) \\ \cos \left( \frac{3t \pi}{24} \right) \end{array} \right) \left( \begin{array}{c} \theta_{14} \\ \theta_{15} \end{array} \right) \right),
$$

where $W(t)$ is an indicator function which equals 1 if the $t$ lies on a weekend. Since the weather data is only available for every hour, we consider the functions $P(t)$ and $T(t)$ as piece-wise constant. Note that the essential assumption of continuity of the link function is not violated because we only require continuity in the parameter. The constant $L_n(t)$ is bounded because temperature and precipitation as well as the other functions are bounded. The Assumption (VX) is fulfilled because the vertex-labels, that is, the bike station IDs, were virtually randomly assigned to the bike stations. The continuity of $t \mapsto \mu_n(t; \beta_0)$ required in Assumption (C) is more of an issue. However, in the limit this is not a problem because the continuity is used for kernel approximations, thus, these approximations are only problematic at the discontinuity. Since we integrate over the entire observation period, these approximation errors on short intervals are not of big importance.

We consider covariates on pair-wise level based on distances. Let $d_{i,j}$ denote the logarithm of the biking time in minutes from station $i$ to station $j$ as returned by Google maps. Then, we consider a simple bi-variate covariate vector

$$
X_{n,i,j} = \left( \frac{d_{i,j}}{a_{i,j}} \right).
$$

As a link function we consider $\Psi(x; \beta) = \exp(x' \beta)$ and hence we have a Cox proportional hazards model. We compute the estimates $\beta_n$ and $(\hat{\theta}_n, \hat{\beta}_n)$ as it was described in Section 4.1 based on the first week, that is, from May 13 to May 19. The estimates $\beta_n$ and $\beta_n$ are almost identical and have roughly the values $(0.219, -0.147)'$. The negative sign of the parameter for $d_{i,j}^2$ shows that the intensity (as a function of the distance) has a maximum. This is plausible because people will likely not use the bikes for very short or very long tours. Now we compute the nonparametric Nelson-Aalen estimator of the baseline intensity based on the second week and compare the estimate with the parametric estimator. Here we have chosen the bandwidth to be 30 min. This has been obtained by eye-balling. The assumption of the independence of the two weeks is plausible because we expect that people do not base their biking decisions on the past week other than possibly indirectly through the weather which we control for.

Both estimators are shown in Figure 3. We can see that in particular the drop in bike rides on Tuesday (Day 3) (possibly due to the rain) is well captured by the parametric model. Also the difference between weekends and working days is well visible in both estimators. However, there appears to be an overestimation of the activity on the weekends by the parametric estimator. We apply our test which is based on a modified $L^2$-distance (using smoothing and weighting) as described in (5). By using the weight function $w$ we can restrict the test to working days, weekends or the whole week. In all three cases, the test based on $T_n$, centralised and scaled according to Theorem 3.1, rejects the hypothesis of a baseline that is entirely driven by global covariates. The observed $p$-values are $1 - \Phi(364) \approx 0$ (whole week), $1 - \Phi(627) \approx 0$ (working days), and $1 - \Phi(169) \approx 0$ (weekend). To investigate the deviation further, it is natural to compare the deviation between parametric and nonparametric

![Density of Bike Rides](image)

**Figure 1.** Kernel density estimate for bike rides over a period of two weeks in 2018 (May 13–May 26).
Figure 2. Temperature and precipitation in Washington D.C.

Figure 3. Parametric (P) and Nonparametric (NP) estimates of the baseline intensity for the second week.
estimator relative to the parametric estimator in the $L^2$-norm
\[
\frac{\sqrt{\int_0^T \left( \hat{\alpha}_n(t; \hat{\beta}_n) - \alpha(\hat{\theta}_n, t) \right)^2 dt}}{\sqrt{\int_0^T \alpha(\hat{\theta}_n, t)^2 dt}} \approx 0.32,
\]
that is, around 32% of the baseline $\alpha_0$ are contributed by the deviation $c_n \Delta \alpha$ and the remaining 68% are explained by the global covariates. Our Theorem 3.1 and the Remark 3.3 allow us to make statements like this, for example, 95% confidence level, as follows. According to Remark 3.3 we can compute that, with probability 95%, $\|K_h \star (c_n \Delta \alpha)\|_w \geq 0.22$. Dividing this number by $\|K_h \star \hat{\alpha}_n(\cdot; \hat{\beta}_n)\|_w$ yields a contribution of around 29.4% or more by $c_n \Delta \alpha$. Put differently, at most 70.6% of $\alpha_0$ are explained through the global covariates. Using the weight function $w$ to restrict to working days only, yields a contribution of $c_n \Delta \alpha$ of 28.8% to $\alpha_0$. Repeating the same exercise on weekends, yields a percentage of 31.2%. Hence, we see in both cases a relatively large misfit of the parametric model.

### 4.3. Simulation

In this section we will simulate the power function of the test procedure. In order to simulate reasonable data, we fit the following model to the bike data from Section 4.2: As pair-wise covariate $X_{ij}$ we use the number of joint neighbors of $i$ and $j$, that is, the number of directed paths from $i$ to $j$ of length 2. For the global covariates we use the same parametric model as in Section 4.2. This results in estimates $\hat{\theta}_n \in \mathbb{R}^{17}$ and $\hat{\beta}_n \in \mathbb{R}$. The fitted parameters are used in the subsequent simulations as true underlying parameters. In particular, we choose $\beta_0$ equal to $\hat{\beta}_n$. As baseline we use $\alpha_0(t) := \alpha(t) + \rho(\alpha_1(t) - \alpha(t))$, where $\rho \in [0, 1]$ and $\alpha_1(t) = \alpha(\hat{\theta}_n, t)$ using the weather information from Section 4.2. The function $\alpha_1$ changes hourly over a period of two weeks, thus, it takes $14 \cdot 24$ different values. The function $\bar{\alpha}$ is constant and it is the average of $\alpha_1$:

\[
\bar{\alpha}(t) = \log \left( \frac{1}{14 \cdot 24} \int_0^{14 \cdot 24} \alpha_1(s) ds \right)
\]

Hence, for $\rho = 0$ the null hypothesis is

\[ H_0 : \quad \alpha(t) = \exp(\theta) \text{ for some } \theta \in \mathbb{R}, \]

The value of $\rho$ controls the deviation from the null hypothesis. Figure 4 shows $\alpha_0(t)$ for the choices $\rho = 0, 0.05, 0.1, 0.2, 0.5$.

In order to obtain a network randomly changing in time, we adopt the model from Jiang, Li, and Yao (2020) to our situation by allowing the parameters for weekends and working days to be different. We fit this model to the observed data from Section 4.2 and again we use the fitted parameters as model parameters for the simulations. We consider two simulation set-ups similar to Section 4.2: One where the network has $n = 527$ vertices (as in Section 4.2) and one for a smaller network with $n = 300$ vertices. In each instance of the simulation, we generate a period of two weeks and split the period in two segments of one week each. We then use the estimation procedure as explained in Section 4.1.
We simulate 2000 such datasets from this model for the choices \( \rho = 0, 0.05, 0.1, 0.2, 0.5 \) (see Figure 4 for a plot of the baselines). For each choice of \( \rho \) we compute the percentage of rejections of our test when it is calibrated to level \( \alpha = 0.05 \). The result is shown in Figure 5. We see that for both network sizes the test complies with the desired level on the hypothesis \( \rho = 0 \) and that it has power against the alternatives. We show in Section 8 of the supplement the finite sample distributions of the test statistics. It can be seen that the normal approximation is quite good justifying tests based on the asymptotic result.

5. Conclusion

In this article we have proposed a test statistic for parametric network models of global covariates. We use weak dependence assumptions to derive the asymptotic behavior of the test statistic. This can be used to obtain critical values. Moreover, from a mathematical point of view, we have illustrated how weak dependence assumptions on networks can be developed in order to allow the derivation of asymptotic results. The assumptions allow for a certain dependence between neighbors in a network but require that the dependence is weak for vertices which are far apart in the network.

As an illustration we have applied the test to bike share data. We saw that even though the weather and time information used by the model appear to explain the data well, there is still some structure that is not unraveled by the particular model under study. This motivates further study in the direction of modeling interactions on networks. As a direct consequence, an interesting future research question would be to test if the fully nonparametric model from Kreiss, Mammen, and Polonik (2019) offered needed flexibility or if a hybrid model like it was studied here offers enough flexibility. More generally, the link-functions \( \alpha \) and \( \Psi \) can be modeled parametrically or nonparametrically, yielding four possible models. In order to understand the situation in a real-world dataset well, all \( \binom{4}{3} \) possible comparisons are of interest. In addition, a thorough treatment of the bandwidth would be desirable.

Supplementary Materials

supplement: The Supplement contains technical proofs and other material which was left out in the main text (pdf). More precisely, the supplement contains a discussion of the role of Kreiss (2021) and how the Mixing assumptions stated there imply Assumption (LL) in Section 6. In Section 7 we discuss our assumptions in more detail, specifically Section 7.1 contains a Motivation for the Assumption (mDep), further information about hubs is provided in Section 7.2, weaker but more technical assumptions which would also imply Theorem 3.1 are provided in Section 7.3. Section 8 contains additional simulation assumptions. Finally, Section 9 contains the proof of Theorem 3.1. More precisely, prerequisites for the proof of Theorem 3.1 are given in Section 9.1, Section 9.2 states Rebollodo's Martingale Central Limit Theorem for the convenience of the reader, the actual proof of Theorem 3.1 can be found in Section 9.3, and further Lemmas needed for the proof of Theorem 3.1 are stated and proved in Section 9.4.

Acknowledgments

Computations for this work were done (in part) using resources of the Leipzig University Computing Centre. W. Polonik acknowledges partial support by the National Science Foundation under Grant No. DMS-1713108. We thank the associate editor and anonymous referees who helped to improve the paper by providing very helpful comments.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Aalen, O. (1978), “Nonparametric Inference for a Family of Counting Processes,” The Annals of Statistics, 6, 701–726. [459] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), Statistical Models Based on Counting Processes, New York: Springer. [459] Andersen, P. K., and Gill, R. D. (1982), “Cox’s Regression Model for Counting Processes: A Large Sample Study,” The Annals of Statistics, 10, 1100–1120. [459] Butts, C. T. (2008), “A Relational Event Framework for Social Action,” Sociological Methodology, 38, 155–200. [457] Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B, 34, 187–220. [459] Gregory, G. (1977), “Large Sample Theory for U-statistics and Tests of Fit,” The Annals of Statistics, 5, 110–123. [460] Härdle, W., and Mammen, E. (1993), “Comparing Nonparametric versus Parametric Regression Fits,” The Annals of Statistics, 21, 1926–1947. [458,459,461] Jiang, B., Li, J., and Yao, Q. (2020), “Autoregressive Networks,” arXiv 2010.04492. [467] Kreiss, A. (2021), “Correlation Bounds, Mixing and m-dependence Under Random Time-Varying Network Distances with an Application to Cox-Processes,” Bernoulli, 27, 1666–1694. [457,458,462,468] Kreiss, A., Mammen, E., and Polonik, W. (2019), “Nonparametric Inference for Continuous-Time Event Counting and Link-based Dynamic Network Models,” Electronic Journal of Statistics, 13, 2764–2829. [457,458,461,462,468] Leucht, A., and Neumann, M. (2013), “Dependent Wild Bootstrap for Degenerate U- and V-Statistics,” Journal of Multivariate Analysis, 117, 257–280. [460] Martinussen, T., and Scheike, T. H. (2006), Dynamic Regression Models for Survival Data, Statistics for Biology and Health, New York: Springer. [459] Matias, C., Rebalka, T., and Villers, F. (2018), “A Semiparametric Extension of the Stochastic Block Model for Longitudinal Networks,” Biometrika, 105, 665–680. [457] Nelson, W. (1969), “Hazard Plotting for Incomplete Failure Data,” Journal of Quality Technology, 1, 27–52. [459] Perry, P. O., and Wolfe, P. J. (2013), “Point Process Modelling for Directed Interaction Networks,” Journal of the Royal Statistical Society, Series B, 75, 821–849. [457,458,459] Ramlau-Hansen, H. (1983), “Smoothing Counting Process Intensities by Means of Kernel Functions,” The Annals of Statistics, 11, 453–466. [459] Scheike, T. H., and Martinussen, T. (2004), “On Estimation and Tests of Time-Varying Effects in the Proportional Hazards Model,” Scandinavian Journal of Statistics, 31, 51–62. [462] Schiebinger, G., Wainwright, M., and Yu, B. (2015), “The Geometry of Kernelized Clustering,” The Annals of Statistics, 43, 819–846. [461]