Characterizing Wireless Networks by Spatial Correlations

Janne Riihijärvi, *Student Member, IEEE*, Petri Mähönen, *Senior Member, IEEE,* and Michael Rübsamen

Abstract— We show that n-point correlation functions are powerful tools to characterize the structure of wireless networks. We argue that they convey the most interesting correlation structures in node locations, while being significantly easier and faster to compute than the classical correlation functions of stochastic geometry. As a case study we analyze the spatial structure of the WLAN access point locations in the east and west coasts of the USA. It is found that measured access point locations feature power-law or scale-free behavior in their correlation structures. Node location models commonly used in simulations and analytical calculations are also studied, and found to be unrealistic in this regard.

Index Terms— Interference, location correlations, pair-correlation functions.

I. INTRODUCTION

SPATIAL structure of wireless networks is still rather poorly understood subject. During the last decade our understanding of the topological structure of the Internet has increased tremendously (see, for example, [1], [2] for recent reviews). However, this very interesting work has primarily targeted only *fixed* networks, while much less attention has been placed on the study of the structure of *wireless* networks. Naturally, similar graph-based or *relational* techniques are applicable to both, but whereas for fixed networks the relational topology is arguably a fundamental property, in the wireless case it is purely a derived quantity, dependent on the spatial structure of the network.

The purpose of this letter is to present a foundation for "first-principles" characterization of spatial structure of wireless networks in the form of spatial n-point correlation functions. From the precedent set forth by the graph-based approach mentioned above, we can expect two types of benefits from establishing a basic theory of spatial network structure. First major application is development of more realistic simulation environments. For the networking research community at large, simulators are the most often used tool for performance evaluation. However, as has been pointed out by several authors (see, for example, [3], [4]) the node location and mobility models offered by present-day simulation environments are not realistic. For mobility models there has been some work done towards both basic theory of motion, and for characterizing real-world mobility [5], [6], [7]. Much more limited effort has been directed towards characterizing node locations themselves in *quantitative* manner. To the best of our knowledge, this letter reports first major step into this direction. In particular, the examples presented here are first

ever detailed characterizations of spatial structure of real large-scale wireless networks.

This work also has impact on theoretical analysis and modeling of communication networks. The usually chosen "null hypothesis" of node locations being sampled from binomial or (stationary) Poisson point process is convenient, but unfortunately rather unrealistic. The interesting research problem is to come up with improved models that are still usable for analytical calculations.

The rest of this letter is organized as follows. In Section II we describe briefly how node locations can be interpreted as a realization of a point process. We also review the basic spatial statistics for characterizing point processes before discussing practically computable node location correlations in more detail in Section III. We present a case study in Section IV, characterizing the access point locations in the east and west coasts of the USA, before concluding the letter in section V.

II. NODE LOCATIONS AS POINT PROCESSES

As is usual in wireless networking research, we shall adopt the probabilistic viewpoint on studying node locations. We assume that the observed locations are realization of some unknown *point process*, the properties of which we try to estimate by statistical means. We shall begin by giving a very brief overview of the theory of point processes, before discussing their statistical description in more detail. We shall mostly be following the excellent monograph of Karr [8].

A *point process* N on a region A is a random counting measure, that is, it assigns a random natural number $N(A_i)$ to any suitably regular subset A_i of A. Naturally, N can be thought of a random distribution of indistinguishable points in A, and $N(A_i)$ as the number of points lying in A_i. We can always write $N = \sum_i \varepsilon_{X_i}$, where ε_x is the point mass at x (the measure equivalent to Dirac delta distribution), and X_i is an A-valued random variable. In simulations binomial point processes are typically used, obtained by superimposing $N(A)$ uniformly distributed random points in A, or Poisson point processes, which are obtained from binomial ones by letting $N(A)$ be a random variable with Poisson distribution with mean (regular, diffuse) measure $\mu(A)$.

The structure of N can be characterized in terms of other spatial measures derived from N. Most important examples are given by *void probabilities* $z_N(A) := P\{N(A) = 0\}$, and *mean measure* $\mu_N(A) := E\{N(A)\}$. Putting $N^k(dx_1, \ldots, dx_k) := N(dx_1) \cdots N(dx_k)$, we can also define the *moment measures* $\mu_N^k := E\{N^k\}$ and *covariance measure* by $\rho_N(A_1 \times A_2) = \mu_N^2(A_1 \times A_2) - \mu_N(A_1)\mu_N(A_2)$. These are the analogies of moments and covariances of random variables to point processes. For example, for Poisson point process we have $z_N(A) = \exp(-\mu(A))$, $\mu_N(A) = \mu(A)$ and $\rho_N(A_1 \times A_2) = \mu(A_1 \cap A_2)$, as expected.

While we have found the above statistics practical tools for theoretical work, node location correlations appear to be more expressive, and not much more complicated to define. Fundamentally, correlations are characterized by the *remainder* of a point process, which would naively have in region Γ the form

$$Q((x_1, \ldots, x_k), \Gamma) = P\left\{ N - \sum_i \varepsilon_{x_i} \in \Gamma \;\middle|\; N(\{x_i\}) = 1 \right\}$$

if the locations of x_i are known. The problem with this formula is, of course, that the corresponding measures being diffuse, the probabilities used in conditioning are typically zero. Thus either a more refined approach is necessary, or we have to modify the problem statement slightly.

The first choice leads to the theory of Palm distributions, which give rigorous meaning for Q. We, however, choose the second path, and instead of obtaining *absolute* location correlations seek to find correlations in *relative distances*, which leads into definition and use of n-point correlation functions. Reason for this is that in our context, there does not exist any fixed "background" spatial field relative to which we could even study absolute correlations.

III. n-POINT CORRELATION FUNCTIONS

Suppose the nodes in question have overall (area) density ρ_0. Then the *pair-correlation function* $\xi(r)$ measures the *excess probability* of finding a neighboring node at distance r from a given node. We can capture this definition in terms of probability density

$$dP = \rho_0^2 \big(1 + \xi(r)\big) \, dA_1 \, dA_2$$

of finding one node in each of the area elements dA_i. Similar definition can be given for *cross-correlation function* $\xi_{ab}(r)$, which measures the symmetric excess probability of finding a node of class a at a distance r from a node of class b, that is,

$$dP = \rho_a \rho_b \big(1 + \xi_{ab}(r)\big) \, dA_1 \, dA_2.$$

These definitions can, of course, be extended in straightforward manner to the case of n-point correlations by

$$dP = \rho_0^n \big(1 + \xi^{(n)}(r_{1,2}, r_{1,3}, \ldots, r_{n-1,n})\big) \, dA_1 \cdots dA_n,$$

where $r_{i,j}$ denotes the distance from node i to node j.

We shall now tackle the problem of estimating $\xi^{(n)}$ from a collection of observed node locations, following the long line of solutions developed by the astrophysics community [9]. Given a catalogue of n points, we proceed in Monte Carlo fashion and generate a random point distribution R of $m \gg n$ locations by drawing them from a binomial point process (which does not exhibit correlations in the large m limit). In the pair-correlation case we then denote by DD the number of distinct pairs from the catalogue in a given range r, by DR the number of "cross pairs" in the same range multiplied by $n/2m$, and by RR the number of pairs from the random catalogue. These can be written formally in normalized form
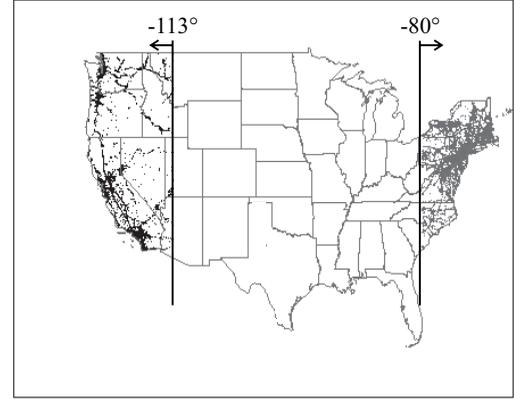

Fig. 1.   The data set of access point locations used in the case study.

as

$$DD(r) = \sum_{x \in D} \sum_{y \in D} [x \neq y] \Phi_r(x, y) / (n(n-1)), \qquad (1)$$

$$DR(r) = \sum_{x \in D} \sum_{y \in R} \Phi_r(x, y) / (nm), \qquad (2)$$

$$RR(r) = \sum_{x \in R} \sum_{y \in R} [x \neq y] \Phi_r(x, y) / (m(m-1)), \qquad (3)$$

where the Iverson bracket $[P]$ equals one if P is true, and is zero otherwise, and $\Phi_r(x, y) := [r \leq d(x, y) \leq r + \Delta]$ is the pair-counting function [10]. Several estimators to $\xi(r)$ can now be written down:

$$\hat{\xi}_1 = DD/RR - 1, \qquad (4)$$

$$\hat{\xi}_2 = DD/DR - 1, \qquad (5)$$

$$\hat{\xi}_3 = DD\,RR/DR^2 - 1, \qquad (6)$$

$$\hat{\xi}_4 = (DD - 2DR + RR)/RR. \qquad (7)$$

These are typically called the Peebles and Hauser (or, simply, natural), Davis and Peebles, Hamilton, and Landy and Szalay (LS) estimators, respectively. By generalising the definitions of $DD(r)$, $DR(r)$ and $RR(r)$ into triplets, quadruplets, and so on, we can write down similar estimators for higher-order correlation functions as well. For example, for reduced three-point correlations (with contributions from $\xi(r)$ removed) we have the Jing & Börner estimator

$$\hat{\xi}^{(3)} = \frac{27RRR^2 \cdot DDD}{DRR^3} - \frac{9RRR \cdot DDR}{DRR^2} + 2. \qquad (8)$$

In the large n, large area limit all of the estimators share the same asymptotic behavior of converging to the correct $\xi(r)$. Nevertheless, they exhibit considerable performance differences in smaller samples, and care should accordingly be applied. In our case study below, we simply used all of the estimators, and checked carefully that all of those give equivalent results in the distance ranges of interest. In addition to these pairwise estimators, several geometric estimators have been proposed. As especially Hamilton and LS estimators have been found very robust in widely varying circumstances, we shall not present these additional estimators here.
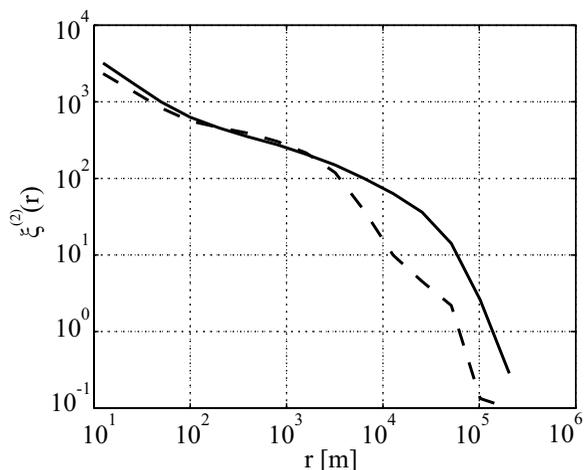
Fig. 2.   The pair correlation functions of the WLAN access point locations for the east (dashed line) and west coast (solid line) data sets.
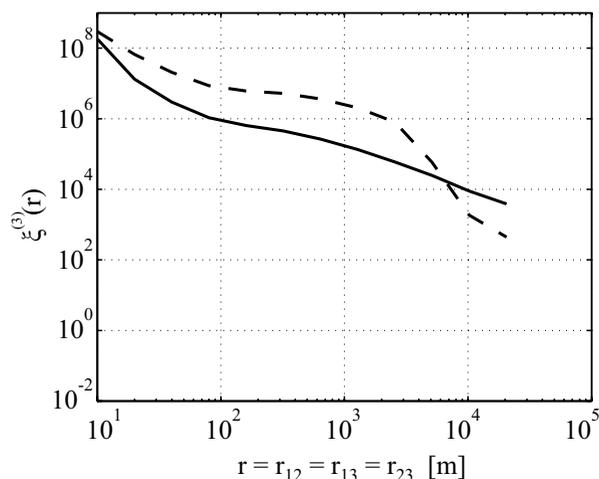

Fig. 3.   The (reduced) three-point correlation functions of the WLAN access point locations for the east (dashed line) and west coast (solid line) data sets. The triplets considered in the estimator form equilateral triangles of length r.

IV. CASE STUDY

Figure 1 shows the point distribution of approximately 10^6 measured Wireless LAN access point locations [11]. The estimators for pair and three-point correlation functions were applied into the data sets, and the corresponding estimates are shown in figures 2 and 3, respectively. As can be seen from the figures the correlation functions follow approximately a (broken) power law. Similar behavior was also observed in four-point correlations, more precise discussion of which is omitted for space reasons. These findings on correlation structures are highly interesting, as most commonly used synthetic location distributions have either vanishing (binomial and Poisson

point processes) or exponentially decaying correlations (most common Neyman-Scott cluster processes, for example). We also conducted tests on other artificial location distributions, such as stationary distributions of various mobility models, but no power-law correlations were observed. There seems to be a fundamental difference in commonly used location models and measured data that certainly warrants further research.

V. CONCLUSIONS

From the above discussions it is obvious that binomial and Poisson point processes or the commonly used clustering processes do not have realistic spatial structure in terms of node location correlations. Accordingly, there is a clear need for devising more realistic point process models for simulation purposes as well as for analytical calculations. As an example, if one is interested in generating point distributions with given density and n-point correlation functions, model of Kerscher can be used [12]. Finally, study of the origins of the observed long-range correlations in Wireless LAN access point locations is an interesting topic for future work.

REFERENCES

[1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, p. 167, 2003.
[2] L. Li, D. Alderson, R. Tanaka, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: definition, properties, and implications (extended version)," Oct 2005. [Online]. Available: http://arxiv.org/abs/cond-mat/0501169
[3] D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," in *Proc. MSWiM '04*. New York: ACM Press, 2004, pp. 78–82.
[4] M. Wellens, M. Petrova, J. Riihijärvi, and P. Mähönen, "Building a better wireless mousetrap: Need for more realism in simulations," in *Proc. WONS 2005*, pp. 150–157.
[5] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. & Mobile Comp.*, vol. 2, no. 5, pp. 483–502, 2002.
[6] F. Bai, N. Sadagopan, and A. Helmy, "The IMPORTANT framework for analyzing the impact of mobility on performance of routing for ad hoc networks," *AdHoc Networks J.*, vol. 1, no. 4, pp. 383–403, Nov. 2003.
[7] F. Bai, N. Sadagopan, B. Krishnamachary, and A. Helmy, "Modeling path duration distributions in MANETs and their impact on routing performance," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 7, pp. 1357–1373, Sept. 2004.
[8] A. F. Karr, *Point Processes and Their Statistical Inference*. Marcel Dekker, 1991.
[9] J. A. Peacock, *Cosmological Physics*. Cambridge University Press, 1999.
[10] M. Kerscher, I. I. Szapudi, and A. S. Szalay, "A comparison of estimators for the two-point correlation function," *Astrophysical J.*, vol. 535, no. 1, pp. L13–L16, May 2000.
[11] WiGLE, http://www.wigle.net/
[12] M. Kerscher, "Constructing, characterizing, and simulating Gaussian and higher-order point distributions," *Physical Review E*, vol. 64, no. 5, p. 56109, 2001.