

## Applicability and limitations of a simple WiFi hotspot model for cities

Michael Seufert, Christian Moldovan, Valentin Burger, Tobias Hoßfeld

### Angaben zur Veröffentlichung / Publication details:

Seufert, Michael, Christian Moldovan, Valentin Burger, and Tobias Hoßfeld. 2017. "Applicability and limitations of a simple WiFi hotspot model for cities." In *13th International Conference on Network and Service Management (CNSM)*, 26-30 November 2017, Tokyo, Japan, 1-7. Piscataway, NJ: IEEE. <https://doi.org/10.23919/cnsm.2017.8255985>.

### Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

**Deutsches Urheberrecht**

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



# Applicability and Limitations of a Simple WiFi Hotspot Model for Cities

Michael Seufert\*, Christian Moldovan<sup>†</sup>, Valentin Burger\*, Tobias Hoßfeld<sup>†</sup>

\*Institute of Computer Science, University of Würzburg, Würzburg, Germany

{seufert | burger}@informatik.uni-wuerzburg.de

<sup>†</sup>Chair of Modeling of Adaptive Systems, University of Duisburg-Essen, Essen, Germany

{christian.moldovan | tobias.hossfeld}@uni-due.de

**Abstract**—Offloading mobile Internet data via WiFi has emerged as an omnipresent trend. WiFi networks are already widely deployed by many private and public institutions (e.g., libraries, cafes, restaurants) but also by commercial services to provide alternative Internet access for their customers and to mitigate the load on mobile networks. Moreover, smart cities start to install WiFi infrastructure for current and future civic services, e.g., based on sensor networks or the Internet of Things. A simple model for the distribution of WiFi hotspots in an urban environment is presented. The hotspot locations are modeled with a uniform distribution of the angle and an exponential distribution of the distance, which is truncated to the city limits. We compare the characteristics of this model in detail to the real distributions. Moreover, we show the applicability and the limitations of this model, and the results suggest that the model can be used in scenarios, which do not require an accurate spatial collocation of the hotspots, such as offloading potential, coverage, or signal strength.

## I. INTRODUCTION

Due to the widespread coverage, end users access Internet services on the run relying mainly on cellular networks. In 2015, the total mobile data traffic reached 3.7 exabytes per month and this monthly volume is expected to surpass 30.6 exabytes in the year 2020 [1]. To cope with these trends, WiFi offloading has gained a lot of attention both by industry and research. Offloading mobile Internet connections to WiFi reduces the load on existing cellular infrastructure, which results in lower infrastructure expenses. Reduced cellular load also leads to increased performance of the network, and thus, to increased customer satisfaction. Moreover, WiFi offloading helps end users to avoid exceeding their data plan volume limitations, and permits the usage of bandwidth demanding applications like video streaming and online gaming also in areas with low mobile coverage.

The deployment of WiFi hotspots is essential for the coverage and the strength of the received signal. Thus, when designing or evaluating services, which rely on the WiFi infrastructure (e.g., mobile traffic management solutions incorporating WiFi offloading [2], Internet of Things services for smart cities [3]), the hotspot locations have to be taken into account. A low signal strength of the WiFi signal results in low throughput, which has an impact on energy consumption

and may not meet the requirements of the application [4], [5], [6]. The distribution of public WiFi hotspots within cities can be obtained from hotspot databases. However, using this approach only the status quo for a given city can be analyzed. To evaluate hypothetical scenarios and the scalability of mechanisms, a generic model would be needed to generate WiFi hotspot distributions for cities of different size, shape, population density, and number of hotspots. Such a model could then facilitate the design and performance evaluation of mechanisms or services, which rely on WiFi infrastructure.

The goal of this work is to extend the work in [7], which presented a simple model for the WiFi hotspot distribution in cities, by investigating its applicability and limitations. Therefore, the WiFi hotspot locations of ten large cities were obtained from a public WiFi database and their characteristics are analyzed. Using a transformation into polar coordinates relative to the city center, we show that the hotspot locations can be modeled with a uniform distribution of the angle and an exponential distribution of the distance. We then compare the characteristics of the real hotspot distributions and the model. We investigate the accuracy of the model for performance evaluation applications, such as offloading potential, coverage, signal strength, interference, handovers, or bandwidth sharing.

This work is structured as follows. Section II describes related work on WiFi hotspot models. Section III shows the applied methodology and Section IV presents the characteristics and model of hotspots distributions in cities. Section VI investigates the applicability of the model for different use cases, and Section VII concludes.

## II. RELATED WORK

WiFi offloading/sharing started in specialized communities (e.g., Fon<sup>1</sup>), but public WiFi is now widely available as both free and commercial services. Many cities over the world have comprehensive WiFi coverage in the city centers just by free public WiFi hotspots provided by various cafes, shops, bars, pubs, libraries, public buildings, and government buildings. There are databases providing the locations of these open/public WiFi hotspots. Many of these databases are user based websites with hotspot locations gathered, uploaded, and updated by a huge community (e.g., OpenWiFiSpots<sup>1</sup>).

<sup>1</sup><http://www.fon.com>, <http://www.openwifispots.com/>

Moreover, also telecommunication operators (e.g., BT<sup>2</sup>) deploy own private/closed WiFi infrastructure to offer their users access to an alternative Internet link. A different WiFi sharing concept is BeWifi<sup>2</sup>, which aggregates spare capacity of close access points for demanding users [8], [9], [10]. Freifunk<sup>2</sup> is a decentralized wireless community network, which uses mesh technology to bring up ad hoc WiFi networks between private WiFi routers, to support local private communication.

The spatial distribution of WiFi hotspots is measured with a tracking method in [11]. The results show that highest density of WiFi hotspots corresponds to residential areas. The distribution of WiFi hotspots is naturally related to the population density in the city, since WiFi hotspots are deployed in close to every household, offices, shops or public places. A first model of the population density with exponential decline from the city center was developed in [12]. A survey on studies of urban population density [13] provides an overview of refined models considering, e.g., lower density in the center due to lower residential land use, or polycentric cities [14].

The relation between the spatial structures of planned wireless networks and population densities has been investigated in [15]. The authors find that base stations belonging to different mobile operators often cluster according to population density. In [16], different point process models are used to model the density of cellular networks. A survey on the literature related to stochastic geometry models for modelling cellular networks is provided in [17]. In a recent work [18], the spatial correlation in base station placements of different mobile operators is investigated and analytical expressions for the coverage probability are derived. However, the distribution of WiFi hotspots differs from the distribution of base stations, as the distribution of base stations is planned and engineered. The models further lack of means to generate hotspot distribution for cities of different shape, e.g., due to natural borders of a coastline. In contrast, we develop a simple model for the more natural distribution of public WiFi hotspots within a city and we investigate its applicability in different scenarios.

### III. METHODOLOGY

To investigate the distribution of public WiFi hotspots in cities, geographic information about the hotspot locations is needed. We use the OpenWiFiSpots<sup>1</sup> database to obtain the addresses of public hotspots. The database is a directory of free WiFi hotspots, which is continuously updated by a growing community of users. Considering that the website provides no API to request the data, the hotspots of ten large cities (eight in the United States, two in Europe) were searched manually on the website and the addresses were parsed from the search results. To obtain more general results, we selected cities with a large number of listed hotspots and different layouts (e.g., grid-based cities, ring-based cities) and characteristics. Table I presents some of these characteristics, i.e., the number of gathered hotspots, the total investigated area, and the population of each city. It can be seen that the

TABLE I  
GENERAL INFORMATION ABOUT INVESTIGATED CITIES.

City	Number of hotspots	Total investigated area (in $km^2$ )	Population (in thousands)
Austin	220	220	843
Berlin	110	250	3502
Boston	193	173	637
Brooklyn (NYC)	454	419	2566
Houston	307	306	2161
Los Angeles	199	165	3858
London	668	367	8308
Portland	419	465	603
San Francisco	214	241	826
Seattle	296	202	635

cities widely differ, for example, in terms of population and hotspot density. Note that the obtained hotspot locations are only a sample of a possibly larger number of WiFi hotspots, as some hotspots might not be listed in the database. As the submission of hotspot locations to the database by the users are independent random processes, we conclude that the obtained sample is a random sample of WiFi hotspots, which will not affect the modeled distributions.

To transform the addresses to geographic coordinates (latitude  $\varphi$ , longitude  $\lambda$ ), the MapQuest<sup>3</sup> geocoding API was used. Moreover, the city center  $(\varphi_c, \lambda_c)$  was computed via a centroid calculation on the WiFi hotspot locations using the k-means algorithm. The hotspot distribution can now be analyzed relative to the city center, which allows more general statements for each city. Mathematically, the geographic coordinates of the WiFi hotspots were transformed into a polar coordinate system, which had the city center as reference point and north as reference direction. Using basic results from spherical trigonometry, coordinates  $(\varphi, \lambda)$  of each WiFi hotspot can be expressed in terms of polar coordinates  $(d, \theta)$  with the spherical distance  $d$  from the city center and angle  $\theta$  towards the reference direction. In Eq. 2, the spherical distance between the coordinates  $(\varphi_c, \lambda_c)$  and  $(\varphi, \lambda)$  (in radians) is computed by using the haversine formula (term  $a$  from Eq. 1) and the mean radius of the Earth  $r_E$ . In Eq. 3, the angle between  $(\varphi, \lambda)$  and the reference direction is calculated. Both computations use the  $atan2$  function<sup>4</sup>, a two argument version of the arctangent function implemented by many programming languages. Note that negative angles of  $\theta$  point counterclockwise from north, whereas positive angles point clockwise from north.

$$a = \sin^2\left(\frac{\varphi - \varphi_c}{2}\right) + \cos \varphi \cdot \cos \varphi_c \cdot \sin^2\left(\frac{\lambda - \lambda_c}{2}\right) \quad (1)$$

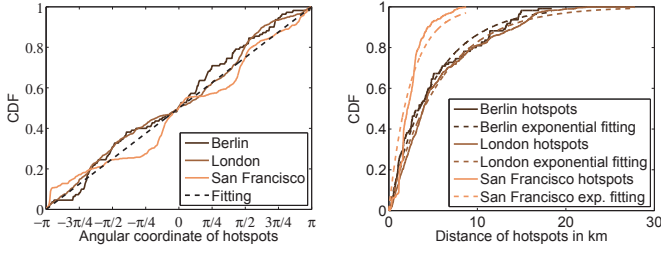
$$d = 2 \cdot r_E \cdot \arctan2(\sqrt{a}, \sqrt{1-a}) \quad (2)$$

$$\theta = \arctan2(\sin(\lambda - \lambda_c) \cdot \cos \varphi, \cos \varphi_c \cdot \sin \varphi - \sin \varphi_c \cdot \cos \varphi \cdot \cos(\lambda - \lambda_c)) \quad (3)$$

<sup>2</sup><http://www.btwifi.co.uk/>, <http://www.bewifi.es/>, <https://freifunk.net/en/>

<sup>3</sup><http://developer.mapquest.com/>

<sup>4</sup><http://www.mathworks.com/help/matlab/ref/atan2.html>



(a) Angular distribution of hotspot data. (b) Distance distribution of hotspot data.

Fig. 1. Fitting of angular and distance distributions.

#### IV. CITY-TRUNCATED EXPONENTIAL MODEL

As described in [7], we investigate the hotspot distribution in terms of the distance and angle of the polar coordinates with respect to the city center in order to formulate a model. We show that a decent approximation can be reached with a uniform distribution for the angle and an exponential distribution for the distance of hotspots.

##### A. Angular Distribution

In Figure 1a, the cumulative distribution functions (CDF) of the angular coordinates is compared for the hotspot data of the cities of Berlin, London, and San Francisco. The angular distribution of the hotspots (solid) shows a high similarity to a uniform distribution  $F(x) = \frac{x+\pi}{2\pi}$ ,  $x \in [-\pi, \pi)$  (black dashed). We observed that the distribution is similar for each of the ten cities and with minor deviations due to city-specific geographic conditions like water areas or parks, which caused hotspot-free spaces at the corresponding angles.

To quantify the goodness of fitting with a uniform distribution, we apply two standard methods for comparing distributions, namely, the maximum absolute error, i.e., the Kolmogorov-Smirnov statistic  $D$ , and the mean absolute error ( $mae$ ). These metrics indicate how far the model is from reality at most ( $D$ ) and on average ( $mae$ ), respectively. In Table II, it can be seen that all fittings have a rather high  $D$ , which shows that the angular distributions are not perfectly uniform due to particular geographic characteristics of the different cities. For example, the shape of the city of Austin contributed to a slightly elliptic hotspot distribution causing the highest  $D$  value. However, the  $mae$  values show low values, which indicate that the angular distribution of hotspots in a city can nevertheless be well approximated by a uniform distribution.

##### B. Distance Distribution

In Figure 1b, the cumulative distribution functions of hotspots (solid) are shown, i.e., the relative frequency of hotspots having a distance to the city center smaller than  $d$ . In this case, a high similarity to an exponential distribution can be observed. Estimating the mean  $\mu$  of the exponential distribution  $F(x, \mu) = 1 - \exp(-\frac{x}{\mu})$ ,  $x \geq 0$  (dashed) from the hotspot data (cf. second column of Table III) with a maximum likelihood estimator, a good approximation is visible.

TABLE II  
MAXIMUM ( $D$ ) AND MEAN ( $mae$ ) ABSOLUTE ERROR FOR UNIFORM FITTINGS OF ANGULAR DISTRIBUTION.

City	$D$	$mae$
Austin	0.1619	0.0677
Berlin	0.0841	0.0401
Boston	0.0809	0.0274
Brooklyn (NYC)	0.1142	0.0537
Houston	0.1023	0.0412
Los Angeles	0.0757	0.0288
London	0.0855	0.0308
Portland	0.0713	0.0244
San Francisco	0.1036	0.0370
Seattle	0.0887	0.0353

TABLE III  
MEAN ( $\mu$ ),  $D$ , AND  $mae$  FOR EXPONENTIAL FITTINGS OF DISTANCE DISTRIBUTION.

City	$\mu$	$D$	$mae$
Austin	3.2041	0.1262	0.0545
Berlin	5.0306	0.0661	0.0270
Boston	2.5224	0.1884	0.0476
Brooklyn (NYC)	5.5942	0.2057	0.0383
Houston	6.7299	0.1645	0.0335
Los Angeles	2.7116	0.1079	0.0542
London	5.7011	0.0488	0.0185
Portland	3.8773	0.1186	0.0339
San Francisco	2.5045	0.2732	0.0710
Seattle	2.9365	0.1136	0.0464

In Table III, the  $D$  values illustrate that the distributions are not perfectly exponential. For example, the highest  $D$  value in San Francisco is caused by the high hotspot density along the northeast waterfront, which cannot be accurately reproduced by an exponential distribution. Again for all cities, the generally low  $mae$  values indicate that yet a good approximation can be reached. It is also noteworthy that the exponential fitting works well for cities of different sizes, although small cities are more prone to inaccuracies due to geographical peculiarities.

##### C. Generation of a Hotspot Distribution for a Generic City

To create hotspot distributions with these characteristics, the coordinates of the city center ( $\varphi_c, \lambda_c$ ) (latitude/longitude) have to be determined first. Then, random hotspot locations will be computed in polar coordinates by generating angle  $\theta$  and distance  $d$ . As observed above,  $\mu$  controls the expansion of the area covered with hotspots. Note that the radius of a circle around the city center including  $p\%$  of the hotspot locations, can be computed as  $\frac{\log(1-\frac{p}{100})}{-\mu}$  ( $p$ th percentile).

In order to create a hotspot distribution with the characteristics observed in the section above, inverse transform sampling can be used. Therefore, for each hotspot location  $(d, \theta)$ , two random numbers  $u, v$  have to be drawn uniformly from the unit interval  $[0, 1)$ . Then, the hotspot distance  $d$  to the city center can be obtained by  $d = \frac{-\log(1-u)}{\mu}$ . The corresponding hotspot angle  $\theta$  can be computed from the second random number  $v$  as  $\theta = 2\pi \cdot v$ .

Eq. 4 and 5 use the trigonometrical functions to transform the polar coordinates  $(d, \theta)$  back to latitude/longitude coordinates  $(\varphi, \lambda)$  (in radians) taking into account the city center



$(\varphi_c, \lambda_c)$  and the spherical Earth with radius  $r_E$ :

$$\varphi = \arcsin\left(\sin \varphi_c \cdot \cos \frac{d}{r_E} + \cos \varphi_c \cdot \sin \frac{d}{r_E} \cdot \cos \theta\right) \quad (4)$$

$$\lambda = \lambda_c + \text{atan2}\left(\sin \theta \cdot \sin \frac{d}{r_E} \cdot \cos \varphi_c, \cos \frac{d}{r_E} - \sin \varphi_c \cdot \sin \varphi\right) \quad (5)$$

The limitation of this naive approach is that a circular and possibly unlimited area will be covered with hotspots. Thus, additionally an accept-reject method will be applied, accepting only hotspot locations within the city limits. This can be used to create a hotspot distribution for a city with a given shape (or any arbitrary area), however, the rejection sampling leads to a truncated distribution with different characteristics than the modeled distribution. To approximate the city limits, the convex hull of hotspot locations will be used in this work.

We illustrate this effect for the city of San Francisco in Figure 2. The hotspot distributions have the same number of hotspot locations and were generated using the fittings presented in Table III, both with the naive approach, and with a rejection of hotspots outside the convex hull of the real hotspot locations (black). It can be seen that the naive approach would place many hotspots outside the city limits (red locations), which are rejected. In the following section, we will investigate the characteristics of the model and the real hotspots in detail.

## V. COMPARISON OF MODEL AND HOTSPOT CHARACTERISTICS

In the model presented above, simple fittings are used to approximate the real distributions. Additionally, the truncation introduced by the rejection process has a potential impact on the characteristics of the generated distribution. Thus, this section investigates if the spatial characteristics of the real hotspot locations can be well replicated by the model.

Figure 3a shows the CDF of the distance of the hotspots to the center for the examples of Berlin (black), London (brown), and San Francisco (orange). The actual distance distributions are depicted as solid lines, the dashed lines show the distributions for the city-truncated exponential model. It can be seen that the rejection results in a truncated distribution (cf. Figure 1b), which manifests in smaller distances.

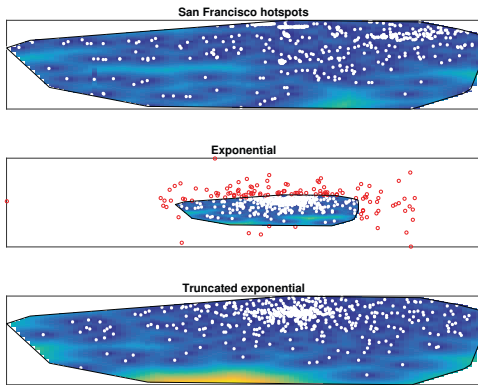
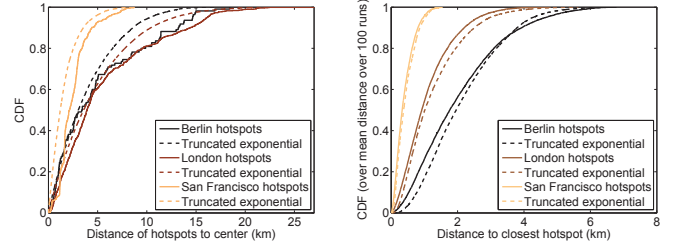


Fig. 2. Impact of truncation on generated hotspot distributions.



(a) Distance of hotspots to city center. (b) Distance to closest hotspot.

Fig. 3. Comparison of hotspot distances for original data and model.

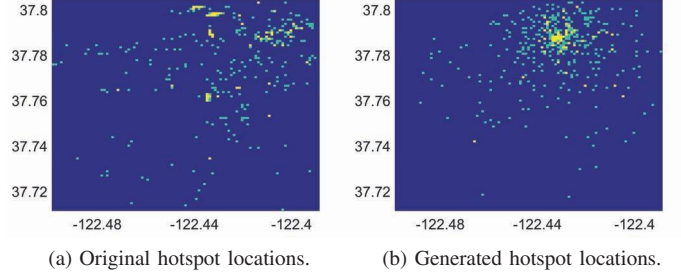


Fig. 4. Comparison of density of hotspots (in terms of number of hotspots in range) for original data and model.

From a practical point of view, the distance from any point to the closest hotspot is important, as it determines the coverage, interference, signal strength, etc. Figure 3b shows the CDF of the distance from a random point within the city to the closest hotspot for the original distributions (solid) and mean distance over 100 generated distributions using the city-truncated exponential model (dashed) for Berlin (black), London (brown), and San Francisco (orange). A high similarity of the generated distributions to the real hotspot distributions is visible, meaning that this characteristics can be well replicated by the model. Later, we will see that this result will propagate to the results for coverage.

In the simple model presented above, the independence of the two dimensions of the polar coordinates is implicitly included, which might not hold for the real data. Therefore, we spatially investigate the hotspot density both for the original hotspot locations in Figure 4a and for a distribution generated by the model in Figure 4b. Both plots show a grid of San Francisco, and the colors indicate the numbers of covering hotspot assuming a WiFi range of 50m<sup>5</sup>. The colors represent the number of hotspots, which can be 0 (blue), 1 (green), or more than 1 (yellow). It can be observed that there are several isolated and scattered hotspots (green), but there are also many hotspots collocated (yellow) in different areas of the cities. Looking in more detail at the collocated hotspots, it can be seen that there are clusters of up to 100 collocated hotspots, for example, in the Marina District in the north. In contrast, the model generates hotspot locations, which are rather regularly concentrated around the center. Also the collocated hotspots can be found near the center implying that the model generates a regular but unrealistic spatial pattern.

<sup>5</sup>Please note that the WiFi range is not of importance, as we are mainly interested in the spatial characteristics in terms of clustering.

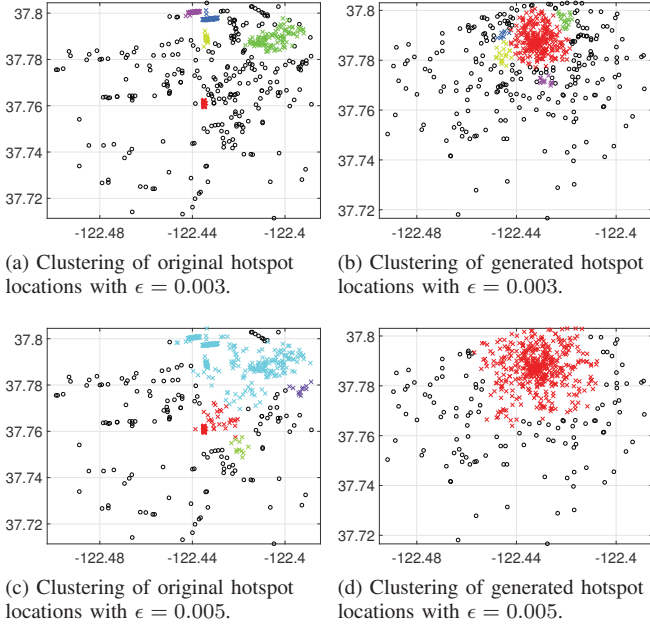


Fig. 5. Clustering with density-based clustering algorithm DBSCAN.

In Figure 5, we take a closer look at the clustering of the hotspots by applying the density-based clustering algorithm DBSCAN [19]. Two density values  $\epsilon = 0.003$  and  $\epsilon = 0.005$  are used (a larger  $\epsilon$  allowing larger distances between hotspots within a cluster), and the minimum number of hotspots of a cluster is set to 10. In Figure 5a, five clusters can be found in the original hotspot locations. There is a large green cluster and four small, highly dense, and spatially separated clusters. When looking at the corresponding plot for the generated distribution (Figure 5b), also five clusters are visible. Again there is one large and four small clusters, however, they are concentrated around the center and are not as dense as the clusters in the original data. If the larger  $\epsilon$  is used, the differences are more striking. The original data has four irregularly shaped and distributed clusters, whereas the generated distribution has a large circular cluster around the center, and thus, does not mimic the clustering of the original hotspot locations.

To quantify the impact of the different clustering, we investigate the spatial autocorrelation of the number of hotspots in range in terms of Moran's  $I$  and Geary's  $C$ . Moran's  $I$  is a measure for global spatial autocorrelation ranging from  $-1$  (regular dispersion) to  $0$  (random pattern) to  $1$  (high clustering). Geary's  $C$  is more sensitive to local spatial autocorrelation and ranges from  $0$  (high clustering) to  $1$  (random pattern) to  $2$  (regular dispersion). Figure 6 shows Moran's  $I$  for different WiFi ranges in the city of San Francisco. Intuitively, as the WiFi range increases, the coverage areas become larger and hotspots are more clustered, and Moran's  $I$  approaches  $1$ . It can be seen that for all ranges, the  $I$  values of the model are very similar to the original data. Table IV presents results for Moran's  $I$  and Geary's  $C$  for all cities assuming a WiFi range of 50m. It can be seen that the spatial autocorrelations of the generated distributions resemble the original distributions even better for the other cities.

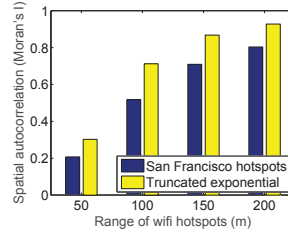


Fig. 6. Spatial autocorrelation of number of hotspots in range.

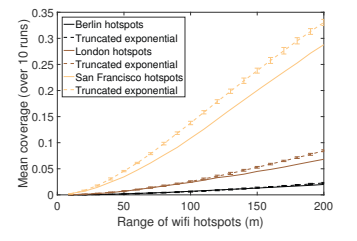


Fig. 7. Mean coverage for different WiFi ranges.

TABLE IV  
SPATIAL AUTOCORRELATION OF HOTSPOT COVERAGE FOR WiFi RANGE OF 50M.

City	$I$ (orig.)	$I$ (model)	$C$ (orig.)	$C$ (model)
Austin	0.0730	0.0783	0.9295	0.9368
Berlin	0.0201	-0.0012	0.9317	1.0164
Boston	0.0918	0.1432	0.9080	0.8718
Brooklyn (NYC)	0.0333	0.0762	0.9732	0.9389
Houston	0.0042	0.0386	1.0110	0.9766
Los Angeles	0.0180	0.0138	0.9680	1.0013
London	0.1420	0.0511	0.8731	0.9640
Portland	0.0720	0.1574	0.9345	0.8577
San Francisco	0.2073	0.3344	0.8066	0.6804
Seattle	0.1301	0.1741	0.8721	0.8409

To sum up, it is obvious that the spatial patterns of the original hotspot distribution cannot be recreated in all details by our simple model. Especially, the clustering of the generated hotspots differs from the original distributions. However, some characteristics can be well approximated by the model. These include the distance to the next hotspot and the spatial autocorrelation of the hotspot coverage.

## VI. APPLICABILITY OF THE MODEL

In this section, three basic scenarios for the applicability of a WiFi hotspot model are discussed. First, we look at the area of WiFi coverage within the city. Second, the number of hotspots in range is investigated. Third, the applicability of the model to investigate WiFi mesh networks is checked.

### A. Coverage Area

The basic scenario for a WiFi hotspot model is to replicate the WiFi coverage, which lays the foundations for evaluations of offloading potential, signal strength, and corresponding applications. The coverage can be computed from the distance to the closest hotspot, cf. Figure 3b. Therefore, for every point it has to be checked if it is in range, i.e., if it has a smaller distance to a hotspot than the WiFi range. Figure 7 depicts the coverage of the example cities Berlin, London, and San Francisco depending on the WiFi hotspot radius, i.e., the WiFi range of each hotspot. Therefore, we generated ten hotspot distributions each and compared the mean coverage percentage and 95% confidence intervals to the original coverage. For all depicted WiFi ranges up to 200m, a high similarity between the real hotspot distribution and the model is visible, which is a consequence of the high similarity in Figure 3b (distance of arbitrary point to closest hotspot). Also increasing WiFi ranges to unrealistic ones up to 1000m would not cause the model to

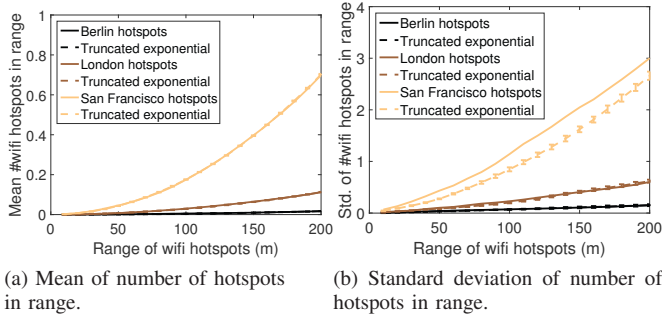


Fig. 8. Comparison of number of hotspots in range.

diverge more from the real data. These results also hold for the other investigated cities. Thus, accurate coverage data can be obtained although the simple model is used. This can be used to investigate the coverage of future cities, for example, to investigate how well WiFi offloading solutions will perform.

### B. Hotspots in Range

The next application of the model is to investigate how many hotspots are in range. This number is the basis for investigations of handovers, interference, bandwidth resource sharing, and corresponding applications. We can compute the number of hotspots in range for a given WiFi radius, cf. the results presented in Figure 4 for a range of 50m. Figure 8a shows the average number of hotspots in range and 95% confidence intervals over 10 runs in the three example cities for different WiFi radii ranging up to 200m. A very high accuracy for the average number can be observed between the model and the real data. However, when investigating the standard deviation of the number of hotspots in range in Figure 8b, large differences can be observed for San Francisco. The mean standard deviation of the model is much smaller than the standard deviation of the original data. This does not come as a surprise, because the differences in the clustering of the hotspots were already visible in Figure 5. Thus, for modeling the number of hotspots in range, the consequences are apparent. It can be seen that this application of the model suffers from its incapability to replicate the spatial structure of hotspots in terms of clustering.

### C. WiFi Mesh Networks

The last application that is considered in this work are WiFi mesh networks. They can rely on direct communication among the WiFi hotspots for privacy reasons (e.g., Freifunk) or to save or aggregate backhaul traffic volume (e.g., BeWiFi). Therefore, we represent the mesh as a graph, in which each hotspot is a node, and two hotspots are connected by an edge if they can communicate, i.e., if they are in WiFi range. The resulting graph representation can then be analyzed by graph metrics. Figure 9a shows a bar plot of the number of connected components in the graphs of the original hotspot locations of San Francisco (blue) and the model (yellow) for different WiFi ranges. Obviously, the number of connected components decreases when the WiFi range is increased. Additionally, it can be seen that the model accounts for higher numbers of

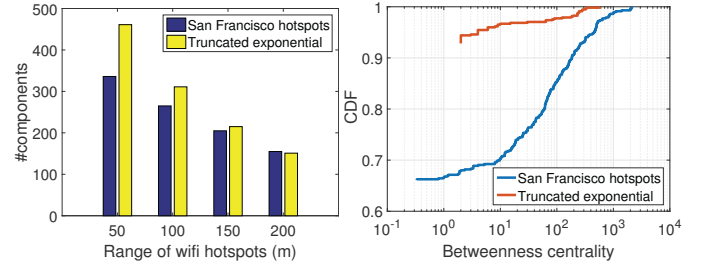


Fig. 9. Metrics for WiFi mesh network graph representation of San Francisco.

connected components for small WiFi ranges. In Figure 9b, we show the CDF of the betweenness centrality in San Francisco assuming a WiFi range of 50m. A high betweenness centrality indicates a high importance of the hotspot for the mesh network as many connections would be routed through this node. It can be seen that both for the original data and the model, many nodes have a betweenness centrality of 0 as they are isolated. However, this number is much higher in the model than in the real distribution. Moreover, it can be seen that the CDF of the model only slowly increases compared to the original data. These are also effects of the different clustering of the original data and the model. The highly centralized hotspot distribution of the model (cf. Figure 5) results in a few highly important nodes. In contrast, the original data is more locally centralized, which causes less isolated nodes and more (locally) important nodes with a high betweenness centrality.

## VII. CONCLUSION

This work investigated the characteristics of the distribution of WiFi hotspot locations in cities. A simple model could be derived, which can be used to create spatial distributions of WiFi hotspots in arbitrary cities. We investigated the characteristics of the generated hotspot locations in detail and compared them to the original data.

It became evident that the spatial patterns of the real hotspot locations cannot be recreated in all details, which leads to inaccuracies with respect to applications considering the collocation or clustering of hotspots, such as handovers, interference, bandwidth sharing, or mesh networks. Also a higher accuracy of fitting, e.g., with Gamma distribution [7], cannot overcome this issue due to the simple assumption that distance and angle of the hotspots are independent.

Nevertheless, we found that for applications that do not require an accurate spatial collocation of the hotspots, but use other characteristics like distance to the closest hotspot, a high accuracy can be achieved by the simple model. For example, offloading potential, coverage, or signal strength in a city could be accurately replicated by the model. Thus, for such applications, the model can be used to generate hotspot distributions to evaluate existing, hypothetical, or future scenarios, for which a real distribution is not available. This can help to design scalable mechanisms, which rely on WiFi infrastructure, and to evaluate their performance for a multitude of different scenarios.

## ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grants HO4770/1-2 and TR257/31-2 (DFG project OekoNet).

## REFERENCES

- [1] Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020,” Cisco, Tech. Rep., 2016.
- [2] M. Seufert, V. Burger, and T. Hoßfeld, “HORST - Home Router Sharing based on Trust,” in *Proceedings of the Workshop on Social-aware Economic Traffic Management for Overlay and Cloud Applications (SETM)*, Zurich, Switzerland, 2013.
- [3] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of Things for Smart Cities,” *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
- [4] V. Burger, M. Seufert, F. Kaup, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, “Impact of WiFi Offloading on Video Streaming QoE in Urban Environments,” in *Proceedings of the IEEE Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*, London, UK, 2015.
- [5] V. Burger, F. Kaup, M. Seufert, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, “Energy Considerations for WiFi Offloading of Video Streaming,” in *Proceedings of the 7th EAI International Conference on Mobile Networks and Management (MONAMI)*, Santander, Spain, 2015.
- [6] M. Seufert, V. Burger, and F. Kaup, “Evaluating the Impact of WiFi Offloading on Mobile Users of HTTP Adaptive Video Streaming,” in *Proceedings of the 5th IEEE International Workshop on Quality of Experience for Multimedia Communications (QoEMC)*, Washington, DC, USA, 2016.
- [7] M. Seufert, T. Griepentrog, and V. Burger, “A Simple WiFi Hotspot Model for Cities,” *IEEE Communications Letters*, vol. 20, no. 2, pp. 384–387, 2016.
- [8] S. Kandula, K. C.-J. Lin, T. Badirhanli, and D. Katabi, “FatVAP: Aggregating AP Backhaul Capacity to Maximize Throughput,” in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, San Francisco, CA, USA, 2008.
- [9] D. Giustiniano, E. Goma, A. Lopez Toledo, I. Dangerfield, J. Morillo, and P. Rodriguez, “Fair WLAN Backhaul Aggregation,” in *Proceedings of the 16th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Chicago, IL, USA, 2010.
- [10] E. Goma Llauro, K. Papagiannaki, and Y. Grunenberger, “A Method and a System for Bandwidth Aggregation in an Access Point,” 2013, WO Patent App. PCT/EP2012/064,179.
- [11] P. Rouveyrol, P. Raveneau, and M. Cunche, “Large Scale Wi-Fi Tracking Using a Botnet of Wireless Routers,” in *Proceedings of the Workshop on Surveillance & Technology*, Philadelphia, PA, USA, 2015.
- [12] C. Clark, “Urban Population Densities,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 114, no. 4, pp. 490–496, 1951.
- [13] J. F. McDonald, “Econometric Studies of Urban Population Density: A Survey,” *Journal of Urban Economics*, vol. 26, no. 3, pp. 361–385, 1989.
- [14] D. A. Griffith and D. W. Wong, “Modeling Population Density Across Major US Cities: A Polycentric Spatial Regression Approach,” *Journal of Geographical Systems*, vol. 9, no. 1, pp. 53–75, 2007.
- [15] M. Michalopoulou, J. Riihijarvi, and P. Mahonen, “Studying the Relationships between Spatial Structures of Wireless Networks and Population Densities,” in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Miami, FL, USA, 2010.
- [16] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.
- [17] H. ElSawy, E. Hossain, and M. Haenggi, “Stochastic Geometry for Modeling, Analysis, and Design of Multi-tier and Cognitive Cellular Wireless Networks: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.
- [18] J. Kibilda, P. D. Francesco, F. Malandrino, and L. A. DaSilva, “Infrastructure and Spectrum Sharing Trade-offs in Mobile Networks,” in *Proceedings of the IEEE Dynamic Spectrum Access Networks Conference (DySPAN)*, Stockholm, Sweden, 2015.
- [19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, Portland, OR, USA, 1996.