On the Dependence Between Base Stations Deployment and Traffic Spatial Distribution in Cellular Networks

Meng Li[∗†], Zhifeng Zhao[∗†], Yifan Zhou[∗†], Xianfu Chen[‡], and Honggang Zhang[∗†]
[∗]York-Zhejiang Lab for Cognitive Radio and Green Communications
[†]Dept. of Information Science and Electronic Engineering
Zhejiang University, Zheda Road 38, Hangzhou 310027, China
Email: {lixiaomeng, zhaozf, zhouyftt, honggangzhang}@zju.edu.cn
[‡]VTT Technical Research Centre of Finland, Oulu 90570, Finland
Email: xianfu.chen@vtt.fi

***Abstract*—Base stations (BSs) deployment and traffic spatial distribution play crucial roles in network design and resource management. Actually, the BSs deployment and traffic spatial distribution are dually coupled, since BSs are built up to fulfill the traffic demand while data traffic is transmitted to mobile users through BSs. In this paper, basing on large-scale measurement data, we firstly show that both BSs deployment and traffic spatial distribution better follow an α-Stable distribution. Specifically, we study the statistical relationship between BSs density and traffic spatial density, which exhibits strong linear dependence for each cellular network generation. Furthermore, according to the slope parameter embedded in the linear regression model, we make a bold but reasonable assumption about the real network configuration evolution pattern, which is concretely manifested by the real transition from 2G to 3G and even 4G.**

I. Introduction

With the popularity of smart portable devices, the demand of cellular data traffic has been increasing rapidly. The Cisco white paper [1] points out that the overall mobile data traffic by 2019 is expected to be a tenfold increase over 2014. To solve the challenges caused by the explosive growth of data traffic, mobile operators have been setting up serveral million base stations (BSs) [2]. Actually, BSs deployment and traffic spatial distribution are fundamentally important but tough issues throughout the long-term evolution of cellular networks. Thus it is imperative to fully understand BSs and traffic spatial distribution as well as statistical relationship between them.

Due to the diverse geographical environment and human social activities (i.e., clustered behaviors, periodic mobility, etc.), traffic in cellular networks expresses dynamic characteristics in spatial dimension [3]. In early time, spatial distribution of cellular traffic, which is served by each BS in a specific time interval, was studied in [4]. With high non-uniformity across different cells, [5] and [6] adopted log-normal distribution to capture traffic density variability instead of traffic. Moreover, burstiness, long-range dependence (LRD) and heavy-tailed properties of broadband wired network traffic have been discovered, and α-Stable model with the above three features was used in [7], [8]. The latest literature [9], first applied α-Stable distribution to model aggregated traffic traces within BSs in the field of cellular networks. Besides traffic spatial distribution itself, its impact on BSs deployment cannot be ignored. In real network scenarios, the locations of BSs are usually coupled with the requirements of subscribers that often exhibit group users' behavior [10]. Also, BSs deployment has been evolving towards the direction of being constantly adapting to the traffic spatial distribution. Recently, poisson point process (PPP) model based on the theory of stochastic geometry [11] has been employed to study the BSs spatial distribution. Meanwhile, Geyer saturation process [12], Poisson cluster process [13], and two-tier PPP [10] have also been proposed to reflect the clustering property of BSs which keeps pace with the social clustering behaviours. Although the above models could provide mathematical tractability in the networking performance evaluation, they failed to reveal the intrinsic heavy-tailed feature of BSs deployment under the influence of traffic spatial distribution. In consideration of this situation, [14] gave a detailed analysis about the heavy-tailed property of BSs deployment and applied α-Stable distribution to describe it.

Moreover, the increasing deployment of dense small cells and multi-tier networking heterogeneity causes the cellular network topology much more complicated than before. Although there have already existed numerous substantial works about the traffic spatial distribution and BSs deployment, the relevant statistical models derived from the former cellular architecture may be impractical to fully reflect the ongoing network evolution. Therefore, by means of analyzing the intrinsic relationship between BSs density and traffic spatial density, we aim to go beyond the temporary changing of network facility, and obtain a deep-level understanding on the fundamental patterns of cellular network evolution in the long term. Previous work like [15], with less BSs and traffic records, adopted saturation model to describe the correlations between the two quantities (i.e., BSs density and traffic spatial density) in urban areas. This result, however, conflicts with the general awareness that BSs distribution and traffic distribution incline to vary consistently.

TABLE I
THE DATASET INFORMATION.

Attributes	City A		City B	
Network Type	2G cellular network		3G cellular network	
Sample Region	Urban1	Rural1	Urban2	Rural2
Size	25×25 km^2	30×35 km^2	20×25 km^2	40×30 km^2
No.of BSs	1061macrocells 847microcells	312macrocells 86microcells	934microcells	467microcells

In this paper, we reveal that both BSs density and traffic spatial density can be much more precisely fitted by an α-Stable distribution, and they mutually are linearly correlated. This result, differing from the recommendation by [15], shows that BSs density and traffic spatial density have almost identical distribution feature. Furthermore, it is verified that a newly defined parameter (i.e., slope k), indicating the needed BSs per unit spatial traffic, can be used to assess the performance of a planned cellular network architecture and forebode network evolution tendency. Practically, this paper would help the operators design network topology more effectively and economically.

The rest of this paper is organized as follows. In Section II, we present a brief description on real data and mathematical background. Section III focuses on the analysis of the statistical relationship between BSs deployment and traffic spatial distribution. In the end, we summarize the paper in Section IV.

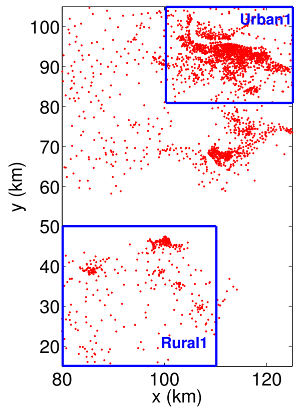


Fig. 1. BS location (red dot) in City A. Two sample regions in blue rectangle are denoted by 'Urban1' and 'Rural1', respectively.

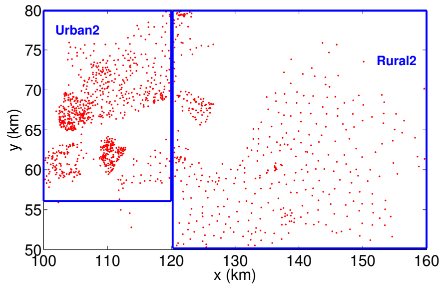


Fig. 2. BS location (red dot) in City B. Two sample regions in blue rectangle are denoted by 'Urban2' and 'Rural2', respectively.

II. BACKGROUND

A. Dataset Description

The measurement data used in this paper is obtained from a commercial mobile operator in China. The dataset, collected from two kinds of networks (i.e., 2G and 3G cellular networks), includes traffic and BSs information of City A and City B. The data traffic is measured in the unit of bytes that each BS transmits to the covered users in one-hour interval. BSs information mainly involves geographic location (i.e., longitude, latitude, etc.) and BS type (i.e., macrocell or microcell). The corresponding details are depicted in Table I.

Specifically, we convert the longitude and latitude values of each BS to X, Y coordinates, and plot the actual geographic location on an 2D coordinate plane as shown in Fig. 1 and Fig. 2. Meanwhile, according to the real city structure, four sample regions named Urban1, Rural1, Urban2 and Rural2 are selected. Obviously, BSs density in rural area is far smaller than that in urban area. Moreover, most BSs in rural area exhibit spatial sparsity. On the contrary, BSs are aggregated densely in urban area, especially in some hotspots.

B. Mathematical Background

α-Stable model with the property of burstiness, LRD and heavy-tailed distribution, manifests itself in the capability to characterize the distribution of normalized sums of a relatively large number of independent identically distributed random variables [16]. Since the probability density function (PDF) is unknown in closed form for most stable distributions, α-Stable distribution is generally specified by its characteristic function.

Definition 1. A random variable X is said to obey α-Stable model if there are parameters $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, and $\mu \in \mathcal{R}$ such that its characteristic function is of the following form:

$$\begin{aligned} \phi(\omega) &= E(\exp j\omega X) \\ &= \exp\left\{-\sigma^{\alpha}|\omega|^{\alpha}\left(1 - j\beta(\mathrm{sgn}(\omega))\Phi\right) + j\mu\omega\right\}, \end{aligned} \tag{1}$$

with Φ is given by

$$\Phi = \begin{cases} \tan\dfrac{\pi\alpha}{2}, \alpha \neq 1; \\ -\dfrac{2}{\pi}\ln|\omega|, \alpha = 1. \end{cases} \tag{2}$$

Here, the function $E(\cdot)$ represents the expectation operation with respect to a random variable. α is called the characteristic exponent and indicates the index of stability, while β is identified as the skewness parameter. α and β together determine the shape of the model. Moreover, σ and μ are called scale and shift parameters, respectively. Specifically, if $\alpha = 2$, α-Stable model reduces to Gaussian distribution.

Usually, it's challenging to prove whether a dataset follows

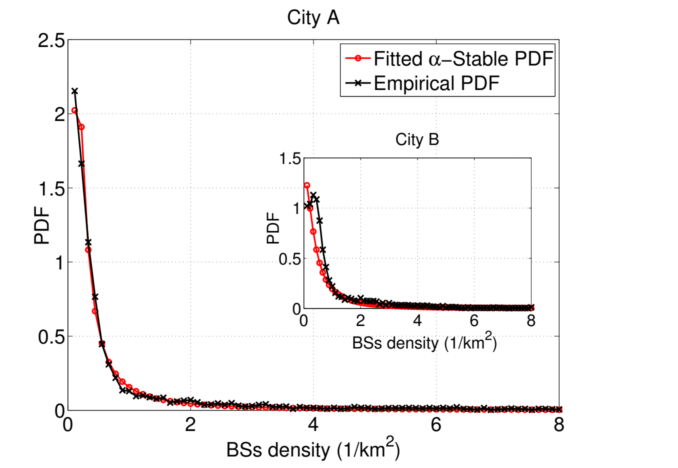


Fig. 3. The results after fitting BSs density to α-Stable distribution in City A and City B, when sampling window size equals 3×3 km^2.

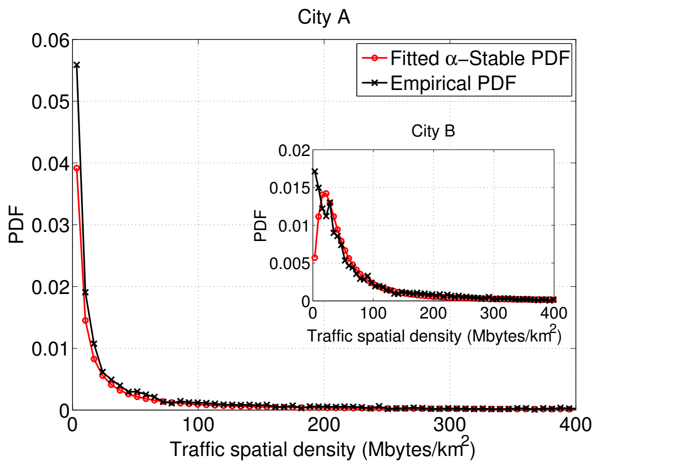


Fig. 4. The results after fitting traffic spatial density to α-Stable distribution in City A and City B, when sampling window size equals 3×3 km^2.

a specific distribution, especially for α-Stable model without a closed-form expression for the PDF. Therefore, when a dataset is said to satisfy α-Stable model, it usually means the dataset is consistent with the hypothetical distribution and the corresponding properties. In other words, the validation needs to firstly estimate the parameters of α-Stable model based on the given dataset, and then compare the real distribution of the dataset with the estimated α-Stable model [8].

TABLE II
THE PARAMETER FITTING RESULTS IN THE α-STABLE MODELS

Object	Attribute	Parameters			
		α(Stability)	β(Skewness)	σ(Scale)	μ(Shift)
BSs density	City A	0.6875	0.8404	0.1363	0.0020
	City B	0.8416	1	0.2274	-0.7463
Traffic spatial density	City A	0.4000	1	5.6098	-0.6105
	City B	0.8726	0.9142	20.9976	-63.9075

III. The Relationship Between BSs Deployment and Traffic Spatial Distribution

A. Spatial Distribution of BSs and Traffic

Humans with similar social behaviours tend to live together, which leads to various traffic hotspots and causes BSs to be deployed densely as clusters in the corresponding areas. [3] pointed out that less than 10% of the subscribers generate 90% of the traffic load while 10% of the base stations carry 50%-60 % of the traffic load, which demonstrates significant traffic imbalance and BSs inhomogeneity in cellular networks. Hence, based on the dataset described in Section II-A, we aim to reveal the inhomogeneity of BSs and traffic distributions.

Firstly, a square sampling window W with size S, is selected randomly. Then, we compute the number of BSs ($N_{S,\mathrm{BS}}$) within this window and that of the aggregated data traffic ($T_{S,\mathrm{TR}}$). Thus, one tuple ($N_{S,\mathrm{BS}}, T_{S,\mathrm{TR}}$) is recorded for each sampling experiment, and the same procedure is repeated 10000 times to obtain enough tuple records. Accordingly, BSs density (λ_{BS}) and traffic spatial density (λ_{TR}) are identified as follows:

$$\begin{aligned} \lambda_{\mathrm{BS}} &= \frac{N_{S,\mathrm{BS}}}{S}, \\ \lambda_{\mathrm{TR}} &= \frac{T_{S,\mathrm{TR}}}{S}. \end{aligned} \tag{3}$$

Considering the real situations that heavy-tailed phenomenon does exist in BSs and traffic spatial distributions, we take α-Stable distribution as the fitting candidate. Based on the statement in Section II-B, the parameters of α-Stable model are firstly estimated and the results are listed in Table II. Afterwards, we use the α-Stable model, produced by the aforementioned estimated parameters, to generate some random variable, and compare the induced PDF with the exact (empirical) one. Therefore, as shown in Fig. 3 and Fig. 4, after fitting an α-Stable distribution to BSs density and traffic spatial density in City A (sampling window size is 3×3 km^2), they both better obey the α-Stable distributions obviously (similar to the findings in [7], [14]). In City B, α-Stable distribution is also applicable.

B. Linear Dependence Between BSs Density and Traffic Spatial Density

Geographic limitations, as well as city structure, lead to the differences of population and traffic demand in diverse regions. Accordingly, the mobile operators adapt their BSs to where the subscribers generate the most traffic. In other words, BSs density is closly related to traffic spatial density. In this section, we will check whether there is any intrinsic correlation between the two quantities.

To ease illustration, Urban1 is taken as an representative example. With the sampling window size being 5×5 km^2, then fitting results are depicted in Fig. 5(b). Evidently, BSs density and traffic spatial density exhibit strong linearity regardless of the BS type. Besides the visual observation, R-square (R^2) value is also adopted as a performance metric to evaluate the goodness of fit. The closer is the R^2 value to

TABLE III
FITTING PARAMETERS OF DIFFERENT GEOGRAPHIC SCENARIOS.

BS type	Sampling Window Size (km^2)	Urban1 k	Urban1 R^2	Rural1 k	Rural1 R^2
macrocell	3×3	0.0226	0.9609	0.0258	0.9887
	5×5	0.0241	0.9890	0.0261	0.9970
	7×7	0.0246	0.9977	0.0262	0.9956
microcell	3×3	0.3916	0.9245	0.3161	0.8357
	5×5	0.4029	0.9503	0.2919	0.8739
	7×7	0.4196	0.9638	0.3060	0.8737

1, the better is the goodness of fit. From Table III, the R^2 value of macrocell and microcell equals 0.9890 and 0.9503, respectively. Therefore, linear regression model is reasonable to characterize the spatial correlation between BSs deployment and traffic spatial distribution, which can be stated as follows:

$$\lambda_{\text{BS}} = k\lambda_{\text{TR}} + t. \tag{4}$$

Here, k is a linear slope value that represents the needed number of BSs per unit spatial traffic.

In order to futher verify the accuracy of using linear regression model, different sampling window size of 3×3 km^2 and 7×7 km^2 are similarly studied, and the corresponding results are illustrated in Fig. 5(a) and Fig. 5(c). Clearly, the sampling window size variation does not violate the linearity. Meanwhile, same tests are carried out in Rural1 and similar conclusions are derived but with different fitting parameters. More detailed numerical information are displayed in Table III.

On one hand, linear regression model keeps better fitting effect no matter the sample region is urban or rural. On the other hand, the key parameter slope k is closely associated with the BS type, without dependence on the sampling window size. These findings indicate that BSs deployment is deeply influenced by the subscribers's demand as well as the corresponding traffic dynamics over the space, and imply that BSs density and traffic spatial density have almost identical heterogeneity feature. Interestingly, it is consistent with the findings in Section III-A that both BSs deployment and traffic spatial pattern demonstrate the same distributed characteristics (i.e., obeying α-Stable distribution), implying the mutual essential interrelation.

C. Cellular Networks Evolution Trend

For City B, BS type is limited to microcell due to the availability of our measurement dataset, but the intrinsic linear dependence still remains the same as depicted in Section III-B. Furthermore, by comparing the fitted parameters in 2G and 3G scenes carefully, we discover that the $k_{\text{2G,microcell}}$ is greater than the $k_{\text{3G,microcell}}$ regardless of region type. The computed results are listed in Table IV. These experimental results demonstrate that an upgraded BS in 3G own more

TABLE IV
SLOPE k OF MICROCELL IN 2G AND 3G CELLULAR NETWORKS, WHEN SAMPLING WINDOW SIZE IS 5×5 KM^2.

k / Network / Region	2G cellular network	3G cellular network
Urban	0.4029 (Urban1)	0.1529 (Urban2)
Rural	0.2919 (Rural1)	0.0772 (Rural2)

capacity and higher transmission rate than that in 2G.

Generally, some technological bottlenecks would be inevitable in cellular networks evolution for each generation. Therefore, new and advanced technologies have been explored to solve the confronted problems, thus achieving success in network upgrading and optimization. Particularly, in view of the difference of slope k in various cellular network scenarios, a reasonable assumption can be stated as follows:

$$k_{2G} > k_{3G} > k_{4G}. \tag{5}$$

This equation indicates that BSs performance of each generation is under continuing improvement together with technology development.

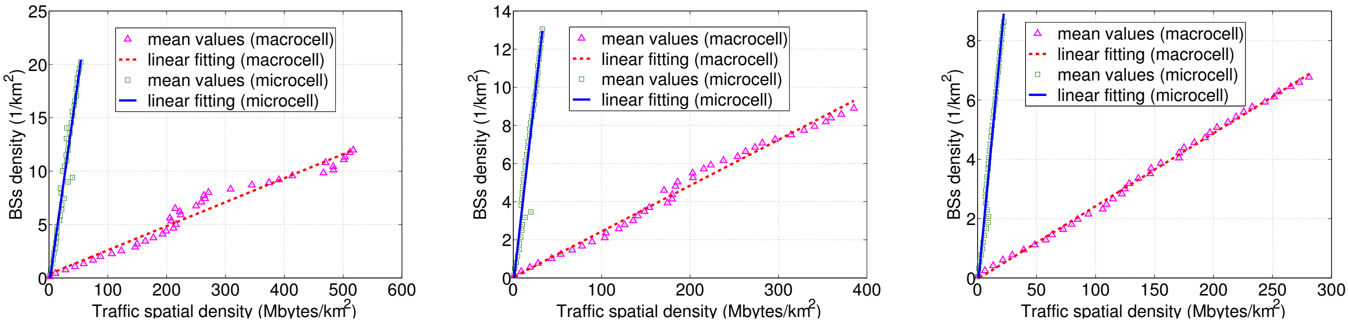
In actual situations, however, with the increase of traffic load, it is impossible for the number of BSs to grow linearly and infinitely, due to the physical and performance constraints of each generation cellular network. Consequently, there should be a certain critical state for each generation cellular network. That is, the available service capability is pre-determined, and if traffic demand increases continuously, the network evolution would go through a network transition (i.e., upgrading from 2G to 3G, then to 4G). In that regard, an explanatory outline about how cellular network architecture evolves is illustrated in Fig. 6. Whether it is a 2G era, 3G era or 4G era, linear dependence between BSs density and traffic spatial density always exists but with different slope k. Surely, the performance improvement of network expects BSs with larger capacity to supply more traffic demand meanwhile requires operators to implement less BSs to serve more subscribers in certain area.

Based on the extensive analyses above, we can reach the following remark.

Remark. *BSs deployment (BSs density) and traffic distribution (traffic spatial density) exhibit a strong linear dependence, which suggests that the heterogeneity feature of the two quantities is almost identical. The slope k in the linear regression model implies the capacity performance of specific BSs and can be adopted as a valuable performance metric in evaluating the long-term evolution of cellular networks.*

IV. CONCLUSION

This paper focuses on the comprehensive study about the statistical relationship between BSs deployment and traffic spatial distribution in cellular networks. For deployed BSs and consumed traffic, α-Stable model characterizes their spatial



(a) 3×3 km^2 (b) 5×5 km^2 (c) 7×7 km^2

Fig. 5. The fitting results of Urban1, when sampling window size varies.

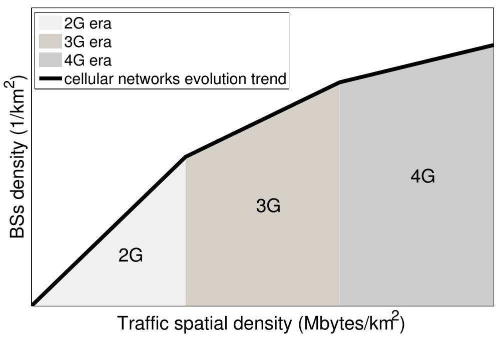


Fig. 6. The real cellular network architecture evolution trend.

distribution much better. Considering the actual situation that BSs placement is deeply induced by traffic dynamics pattern, we reveal the suitability of applying linear regression model to display the statistical relationship between BSs density and traffic spatial density, which is different from the conclusions given by previous references. Additionally, because of the needs to break through various technological bottlenecks, it is unrealistic that the real network configuration will keep unchanged forever. Specifically, the key value slope k would inevitably decrease with cellular network evolution. It can be foreseen that the new findings in this paper will be beneficial for mobile operators on deploying BSs reasonably, thus making network structure optimized and resource utilization improved.

Acknowledgment

This paper is partially supported by the National Basic Research Program of China (973Green, No. 2012CB316000), the Key Project of Chinese Ministry of Education (No. 313053), the Key Technologies R&D Program of China (No. 2012BAH75F01), and the Program for Zhejiang Leading Team of Science and Technology Innovation (No. 2013TD20).

References

[1] Cisco, *Cisco Visual Networking Index: Global Mobile data Traffic Forecast Update, 2014-2019*, Feb. 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

[2] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 4, pp. 524–540, Nov. 2011.

[3] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE INFORCOM 2011*, Shanghai, China, 2011, pp. 882–890.

[4] R. Rathgeber, "Spatial traffic distribution in cellular networks," in *Proc. IEEE VTC 1998*, Ottawa, Ont, 1998, pp. 1994–1998.

[5] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Communications*, vol. 21, no. 1, pp. 80–88, Feb. 2014.

[6] H. Klessig, V. Suryaprakash, O. Blume, A. Fehske, and G. Fettweis, "A framework enabling spatial analysis of mobile traffic hot spots," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 537–540, Oct. 2014.

[7] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[8] G. Xiaohu, Z. Guangxi, and Z. Yaoting, "On the testing for alpha-stable distributions of network traffic," *Computer Communications*, vol. 27, no. 5, pp. 447–457, Mar. 2004.

[9] R. Li, Z. Zhao, C. Qi, X. Zhou, Y. Zhou, and H. Zhang, "Understanding the traffic nature of mobile instantaneous messaging in cellular networks: A revisiting to α-stable models," *IEEE Access*, vol. 3, pp. 1416–1422, Sep. 2015.

[10] J. Zhang, W. Wang, X. Zhang, Y. Huang, Z. Su, and Z. Liu, "Base stations from current mobile cellular networks: Measurement, spatial modeling and analysis," in *Proc. IEEE WCNC 2013*, Shanghai, China, 2013, pp. 1–5.

[11] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[12] D. B. Taylor, H. S. Dhillon, T. D. Novlan, and J. G. Andrews, "Pairwise interaction processes for modeling cellular network topology," in *Proc. IEEE GLOBECOM 2012*, Anaheim, CA, 2012, pp. 4524–4529.

[13] C.-H. Lee, C.-Y. Shih, and Y.-S. Chen, "Stochastic geometry based models for modeling cellular networks in urban areas," *Wireless Networks*, vol. 19, no. 6, pp. 1063–1072, Aug. 2013.

[14] Y. Zhou, R. Li, Z. Zhao, X. Zhou, and H. Zhang, "On the α-stable distribution of base stations in cellular networks," *IEEE Communications Letters*, vol. 19, no. 10, pp. 1750–1753, Oct. 2015.

[15] S. Zhou, D. Lee, B. Leng, X. Zhou, H. Zhang, and Z. Niu, "On the spatial distribution of base stations and its relation to the traffic density in cellular networks," *IEEE Access*, vol. 3, pp. 998–1010, Jul. 2015.

[16] G. Samorodnitsky, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall/CRC, Jun. 1994. [Online]. Available: http://www.amazon.com/Stable-Non-Gaussian-Random-Processes-Stochastic/dp/0412051710