Dynamics of Functional Failures and Recovery in Complex Road Networks Supplemental Material

Xianyuan Zhan and Satish V. Ukkusuri
Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

P. Suresh C. Rao
Lyles School of Civil Engineering and Agronomy Department, Purdue University, West Lafayette, IN 47907, USA

(Dated: October 18, 2017)

Here we elaborate on the traffic state data collection method and the data fitting approach. Additional numerical analyses and the complete data fitting results using the vertex split-recovery model are also presented.

I. DATA COLLECTION

This study uses real world congestion evolution data collected from real-time traffic information provided by Baidu Map. Baidu Map is the most popular on-line digital map service in China. Similar to Google Map, Baidu Map collects network-wide road traffic conditions using crowd-sourced information from users' mobile devices. The travel times from any two given location points can be obtained by querying Baidu Map Route Matrix API, from which the average speed of each road segment in the network can be obtained. The reason that we collected data from Baidu Map in Chinese cities rather than Google Map is because that Baidu Map API allows for a higher free usage limit (100,000 API request per day). The usage limit for Google Map API is only 2,500 request per day, which is not possible for tracking the fine-grained congestion evolution for large urban road networks.

In this study, we collected the congestion evolution data from road networks of two megacities in China: Beijing (road network within 4th ring road, contains 17,148 road segments) and Shanghai (road network within the Middle ring road, contains 18,173 road segments). The road networks of Beijing and Shanghai used in this study, as well as the sample road speed data collected are presented in Fig.1a and 1c. We collected 7 days' network state data from Beijing road network in December, 2015 (12/6-12/8, 12/10-12/12 and 12/14) and 6 days' data from Shanghai in July, 2016 (7/7, 7/9-7/13) using the data crawler running from 6:00 to 24:00 for each day. Collecting travel times for all road segments in the networks takes 40min for Beijing network and about 60min for Shanghai network, which allow us to track the congestion evolution process at an hourly basis. All the road travel time data were converted to road speeds. There are a certain fraction of road segments cannot be resolved by Baidu Map API (less than 30% for Beijing network; less than 40% for Shanghai network), the speeds for these road segments were estimated as the mean value between their upstream and downstream neighboring road segments. The collected road speed data were further compiled into a binary functional state (failed state 0: congested; working state 1: not congested) for each road segment at each time step. In this study, a road segment is identified as congested when its speed is less than 20% of its speed limit. For simplicity, we model the road network as an undirected network. For a road segment that carries bidirectional traffic, we consider the segment in the failed state if traffic of any of the two directions gets congested. The obtained functional states of the network are used to construct the function augmented dual networks to study the congestion evolution process.

The empirical analysis of the real world vertex split-recovery process are primarily based on the data from Beijing road network due to higher data quality (shorter collection cycle and higher fraction of road segments resolved by Baidu Map API). However, we also use the data from Shanghai road network as a supplement to better demonstrate the applicability of the proposed model on different road networks.

All the collected congestion evolution data and the road network data used in this study are publicly available at: `https://github.com/zhanzxy5/VSR-dataset`.

II. DATA FITTING

If we denote $w_1(t) = \eta(t)/\theta$, $w_2 = \tau/\theta$ and $w_3 = \lambda/\theta$, the $\Lambda_1^k(t)$, $\Lambda_2^k(t)$, $\tilde{\Lambda}_1^k(t)$ and $\tilde{\Lambda}_2^k(t)$ can be simplified as

$$\Lambda_1^k(t) = w_1 + w_3\frac{k}{k-1}\phi(t), \quad \Lambda_2^k(t) = \Lambda_1^k(t) + w_2 \quad (1)$$

$$\tilde{\Lambda}_1^k(t) = w_1 + w_3\frac{k}{k-1}\tilde{\phi}(t|k), \quad \tilde{\Lambda}_2^k(t) = \tilde{\Lambda}_1^k(t) + w_2 \quad (2)$$

Thus the stationary solution of the vertex split-recovery model under DBMF approximation can be completely determined by w_1, w_2 and w_3. In addition, $w_1(t)$ can be interpreted as the normalized network loading level, which is a dynamic variable that measures the relative loading level of the entire network; w_2 and w_3 represents the normalized self-contagion and neighbor contagion rate, which are fixed parameters governed by network structure.

We fit the proposed model to empirical data in order to uncover the actual $w_1(t), w_2$ and w_3 values in the

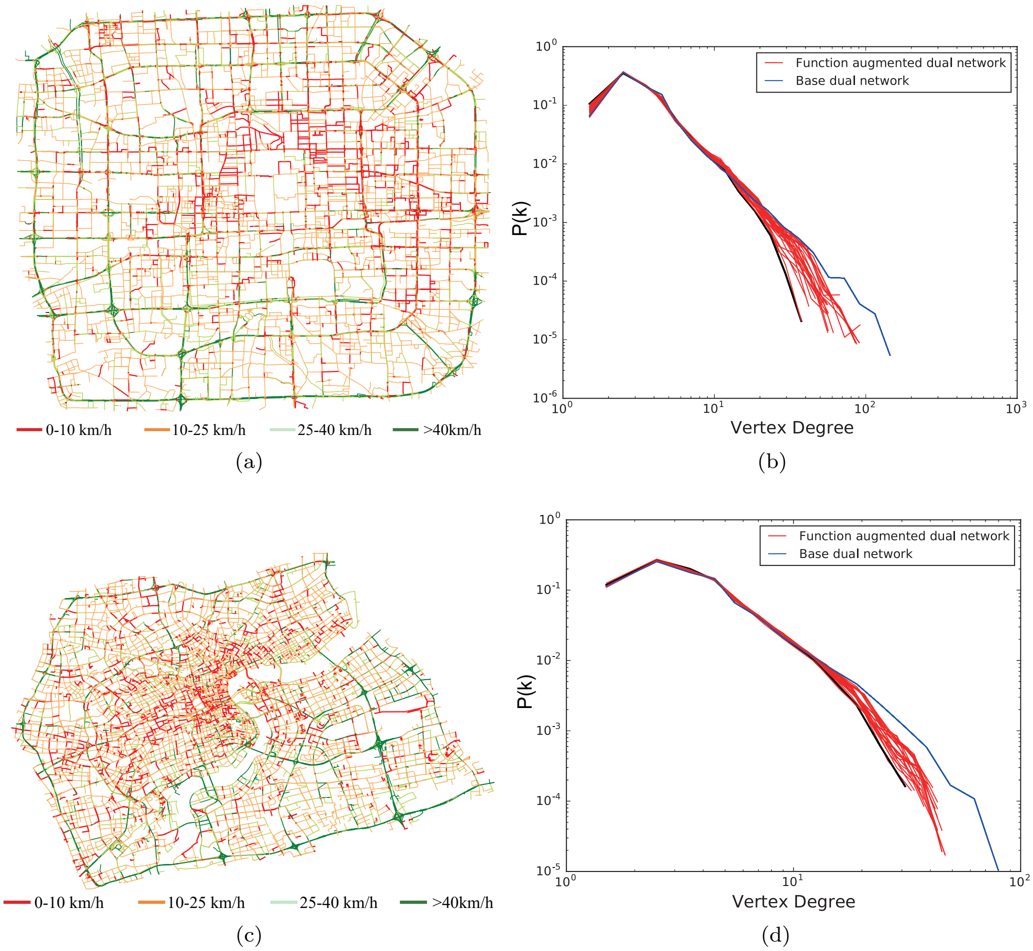

FIG. 1: (a) and (c) are the speed plots for Beijing (2015/12/14 6:00 PM) and Shanghai (2016/7/7 5:30 PM) road networks, which are the most congested hour during the day. (b) and (d) are the evolution of dual degree distributions for function augmented dual network of Beijing and Shanghai road networks during the same day. The black lines in (b) and (d) correspond to congestion scenario showed in (a) and (c). Logarithmic binning is used for better clarity. The degree distribution of the base dual network can be well fitted into a power law distribution $p(k) \sim k^{-\gamma}$ for $k \geq k_{min} = 3, \gamma = 2.58$ (Beijing) and $k \geq k_{min} = 4, \gamma = 2.448$ (Shanghai). In the function augmented dual network, high degree dual vertices are more likely to experience vertex split, causing faster probability decay at the tail. This lead to the deviation of power law distribution, but can be better fitted to power law with exponential cutoff distribution ($p(k) \sim k^{-\gamma} e^{-\kappa k}$).

test networks. Let $P_T^*(k)$ and $Q_T(k)$ be the model predicted and empirical dual degree distribution for time period T. We minimize the overall statistical divergence of the two distributions over the entire day, defined as $min_{w_1(t),w_2,w_3} \sum_T J(P_T^*||Q_T)$, where $J(P_T^*||Q_T)$ is the Jensen-Shannon divergence (JSD) [1] between distribution P_T^* and Q_T, mathematically:

$$J(P_T^*||Q_T) = \frac{1}{2} D(P_T^*||M_T) + \frac{1}{2} D(Q_T||M_T) \qquad (3)$$

where $M_T = \frac{1}{2}(P_T^* + Q_T)$ and $D(P||Q)$ is the Kullback-Leibler (K-L) divergence between distribution P and Q, which can be evaluated as $D(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)}$. JSD is a popular measure for evaluating the dissimilarity between two probability distributions and widely used in statistics and information theory. It has the advantage of being symmetric and can be perceived as the smoothed version of K-L divergence.

In addition to JSD, we also tested using Kolmogorov-Smirnov (K-S) statistics [2] for the data fitting. However, as K-S statistics only measures the supremum of the difference between two cumulative probability functions (CDF), it is relatively insensitive to differences between distributions at the extreme limits of the range of the random variable. As in these limits, the CDFs necessarily tend to zero and one [3]. This feature makes K-S statistics less robust as compared to JSD in our empirical tests. Consequently, we use JSD in the fitting of vertex split-recovery model to empirical data.

III. ADDITIONAL NUMERICAL RESULTS

A. Empirical observations of vertex splits and recoveries

Fig.2 presents the empirical observations of the fraction of roads (dual vertices) with congested segments (vertex splits) for a representative weekday (Thursday, 2015/12/10) and a weekend (Sunday, 2015/12/6) of Beijing road network. The fraction of dual vertices with vertex splits roughly coincide with the variation of demand levels in the city. As for weekday, very high peaks are observed around 7am and 6pm, which happens to be the morning and evening peak; while on weekend, the peak corresponds to the morning peak disappears. It indicates the vertex splitting process is a time dependent process that strongly correlate with the current network demand level.

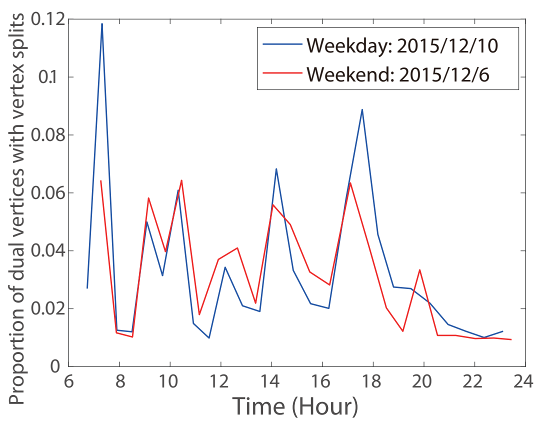

FIG. 2: Empirical observations of the fraction of dual vertices with vertex splits in Beijing road network.

B. Impact of degree correlation in DBMF approximation

As discussed in the main article, there is minor difference in the fitted results of the dual vertex degree distributions between the cases with no degree correlation and general degree correlation under DBMF approximation. This is because that the consideration of general degree correlations only improves the approximation of the neighbor-contagion process. The empirically observed small scale of neighbor-contagion rate w_3 leads to negligible impact on the final dual degree distributions.

However, we do observe that the average probability that a dual vertex has a split neighbor (ϕ^* for no degree correlation case and $\tilde{\phi}(\infty|k)$ for general correlation case) behave differently for different dual vertex degrees. Fig.3 presents the results of the ϕ^* and $\tilde{\phi}(\infty|k)$ values of the stationary solutions under DBMF approximation. Two loading scenarios on Beijing road network are analyzed ($w_1 = 0.01$ for low network loading level and $w_1 = 0.1$ for high network loading level). The resulting dual vertex degree distributions predicted by the vertex split-recovery model for both loading scenarios show highly similar patterns. When comparing the changes in ϕ^* and $\tilde{\phi}(\infty|k)$ values, it is observed that low dual degree vertices tend to have higher average probability to have split neighbors, while high dual degree vertices are the opposite. This could be caused by the dissortativity of the dual road network, which is also reported in other studies [4]. Both the dual road networks of Beijing and Shanghai have negative assortativity coefficient [5, 6] (-0.122 for Beijing and -0.053 for Shanghai), indicating they are disassortatively mixed by degree. This implies star-like network structure that high dual degree vertices tend to connected to low-degree ones. As high dual degree vertices are also more likely to experience vertex split, which results in higher impact on $\tilde{\phi}(\infty|k)$ for low dual degree vertices.

C. Expected maximum dual vertex degree

Fig.4 presents the expected maximum dual vertex degree k^*_{max} of the resulting function augmented dual network under different normalized network loading levels (w_1). k^*_{max} is obtained as $sup\{k : N^*_k \geq 1\}$, where N^*_k is the expected number of dual vertices with degree k in the final function augmented dual network (Eq.19 in the main article).

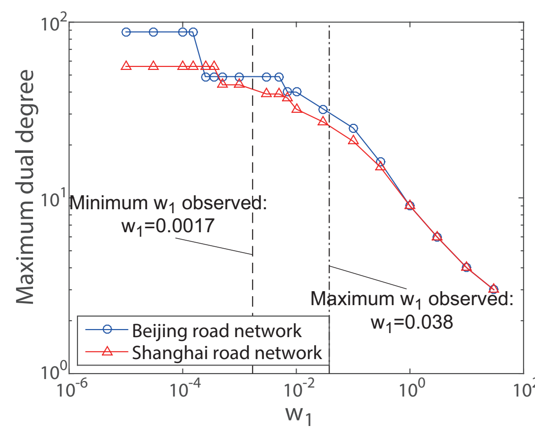

FIG. 4: Impact of w_1 on the expected maximum dual degree of the function augmented dual network. DBMF approximation with no degree correlation is used to obtain the results, other parameters used for computation are: $w_2 = 0.07, w_3 = 0.0005$ for Beijing road network; $w_2 = 0.12, w_3 = 0.0005$ for Shanghai road network.

Both Beijing and Shanghai road networks are investigated. The ring-and-radial structured Beijing road network is more hierarchical compared with the grid like Shanghai road network, which also lead to higher maximum dual vertex degree ($k_{max} = 162$ for Beijing; $k_{max} = 89$ for Shanghai) in its base dual network ($w_1 = 0$). The

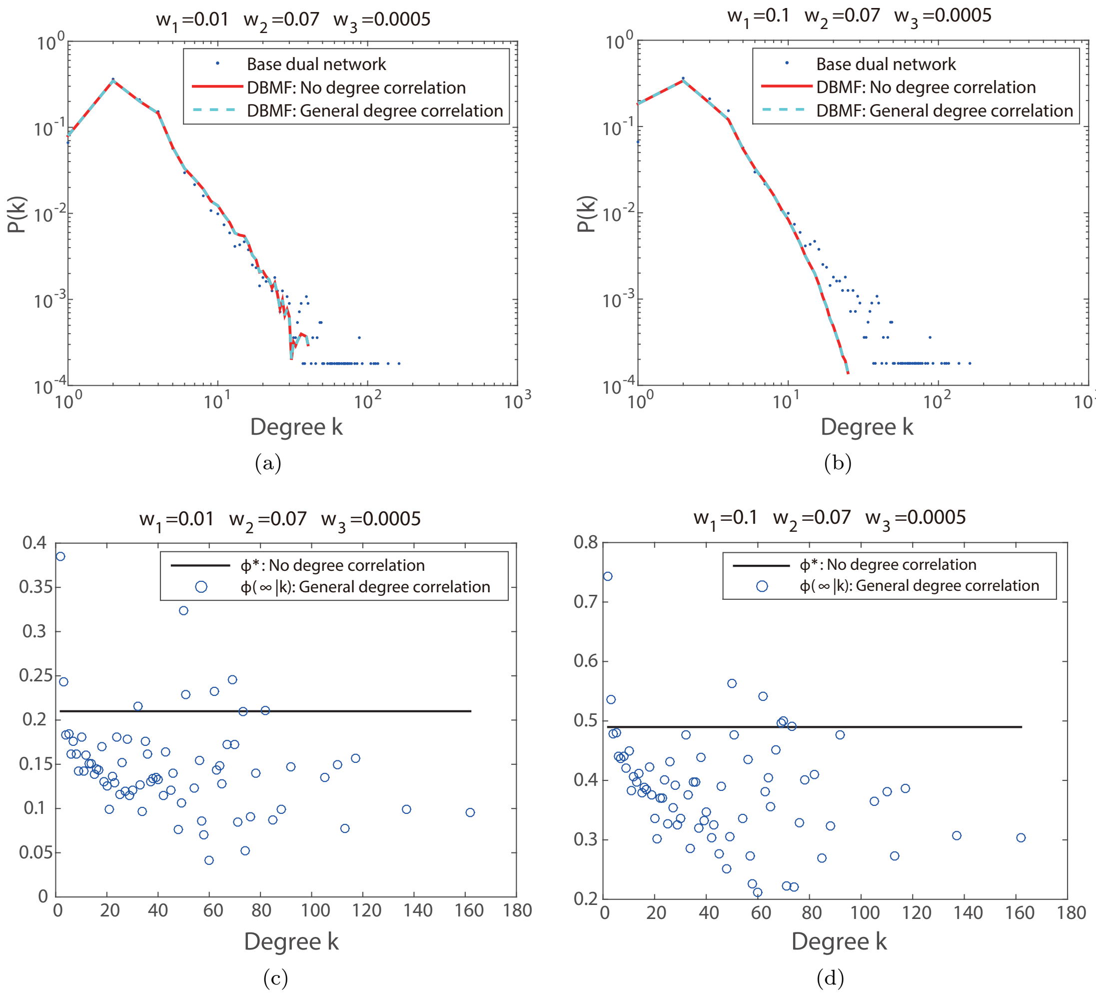

FIG. 3: Comparison of the dual vertex degree distributions ((a) and (b)), and the average probability that a vertex has a split neighbor ϕ ((c) and (d)) under DBMF approximation with no degree correlation and general degree correlation (Beijing road network). (a) and (c) correspond to the scenario with low network loading level ($w_1 = 0.01$); (b) and (d) correspond to the scenario with high network loading level ($w_1 = 0.1$).

higher maximum dual vertex degree k_{max} for Beijing road network also lead to consistently higher expected maximum dual vertex degree in the function augmented dual networks (k^*_{max}) when the traffic loading level is light (small w_1). k^*_{max} can be perceived as the expected maximum degree that a function augmented dual network can maintain under a given network loading level. It can be observed both networks experiences severe high-degree dual vertex loss even under very small network loadings ($w_1 < 10^{-4}$). The stair-like discontinuities in the curve of k^*_{max} is due to the finite network size: there is only a few high-degree dual vertices, split on these dual vertices causes sharp drop in k^*_{max}. After w_1 increases to 1 and becomes larger (self-splitting rate equal or greater than recovery rate), the curve of k^*_{max} for both two networks overlaps. This shows the two road networks with different network configurations seem to have convergent structural characteristics under extreme function failures.

D. Complete fitting results

Fig.9-21 present the complete model fitting results for the collected congestion evolution data (7 days for Beijing road network and 6 days for Shanghai road network) using the data fitting technique described in the Numerical Results section of the main article. Three fitting results are included in each plot, including the vertex split-recovery model under DBMF approximation with no degree correlation and general degree correlation, as well as the power-law with exponential cutoff distribution. The vertex split-recovery model fits the empirical congestion evolution data very well in all the days and time periods. The power-law with exponential cutoff distribution is found to be a good approximation of the results produced by vertex split-recovery model for $k \geq k_{min}$ ($k_{min} = 3$ for Beijing network and $k_{min} = 4$ for Shanghai network). Please refer to Fig.9-21 for details of the fitting results.

E. Comparison with real world data

Fig.5-8 present the complete comparison results of the total number of dual vertices and the vertex splits between the empirical data and the expected values obtained from the vertex split-recovery model. The expected number of vertex splits N^* and the expected total number of vertex splits N^{ES} are computed using Eq.18 and Eq.20 in the main article. The results of N^* and N^{ES} under both DBMF approximation with no degree correlation and general degree correlation are reported. The mean absolute error (MAE) and mean relative error (MRE) are used for evaluation.

It is found that for the total number of dual vertices N^*, the MRE and MAE for Beijing road network are less than 2.5% and 150 respectively. For Shanghai road network, the MRE and MAE are less than 1.5% and 80 respectively. For the total expected number of vertex splits N^{ES}, the MRE and MAE for Beijing road network are less than 16% and 105 respectively. The results of MRE and MAE for Shanghai road network are less than 11% and 72 respectively. On weekends (2015/12/6, 2015/12/12 for Beijing road network; 2016/7/9-10 for Shanghai road network), when the traffic loading level is lower, even smaller MAE and MRE results are achieved for both cities. Considering the relatively low level of MAE and MRE values, the vertex split-recovery model provides a reasonably well approximation of the real world congestion evolution process. Detailed comparison results please refer to Fig.5-8.

[1] J. Lin, IEEE Trans. Inf. Theory **37**, 145 (1991).
[2] F. J. Massey Jr, J. Am. Stat. Assoc. **46**, 68 (1951).
[3] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Rev. **51**, 661 (2009).
[4] S. Porta, P. Crucitti, and V. Latora, Physica A **369**, 853 (2006).
[5] M. E. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[6] M. Newman, *Networks: an introduction* (Oxford university press, 2010).

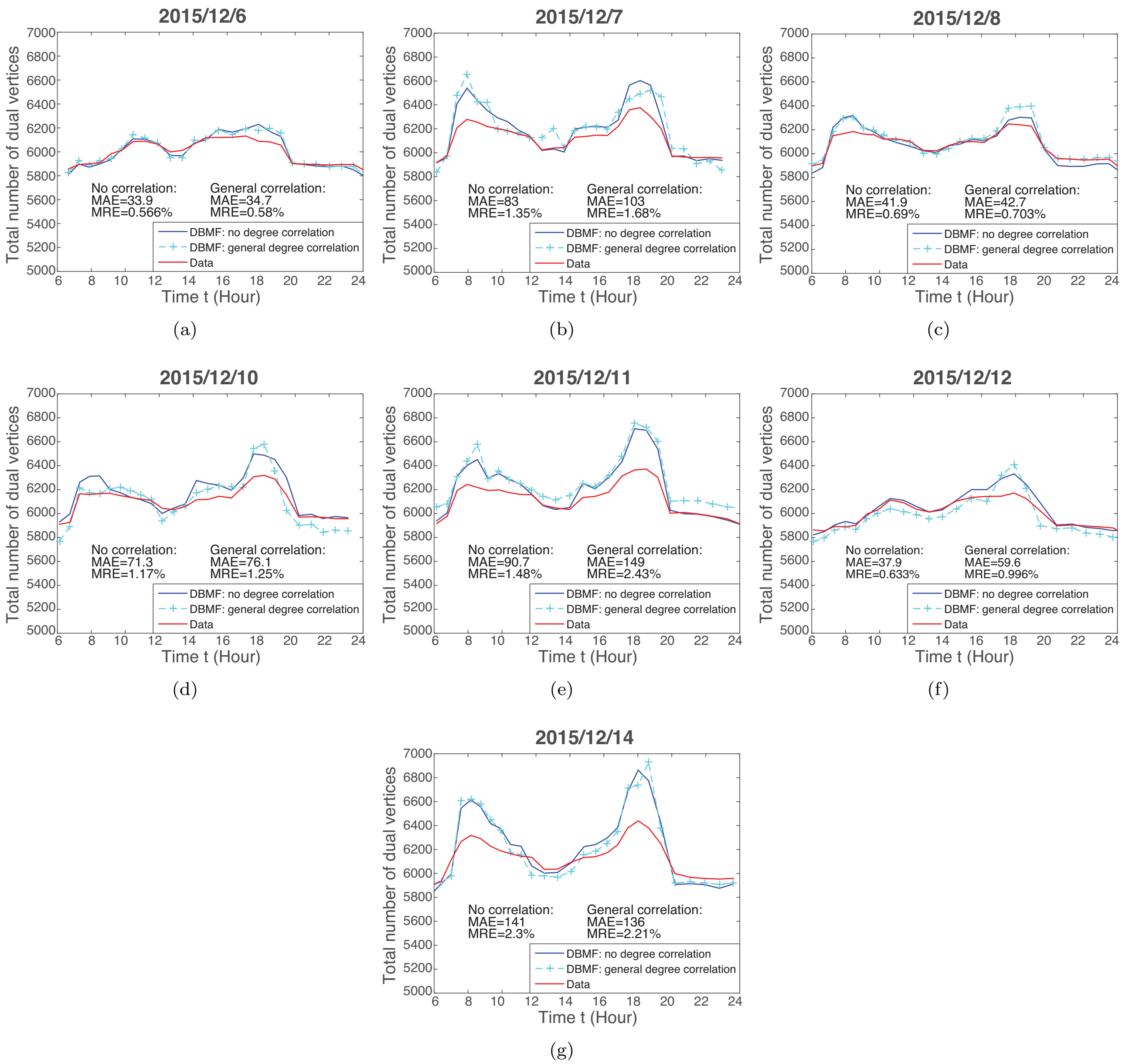

FIG. 5: Comparison of the total number of dual vertices from emprical data (red line) and the expected total number of dual vertices (N^*) obtained from the vertex split recovery model for Beijing road network.

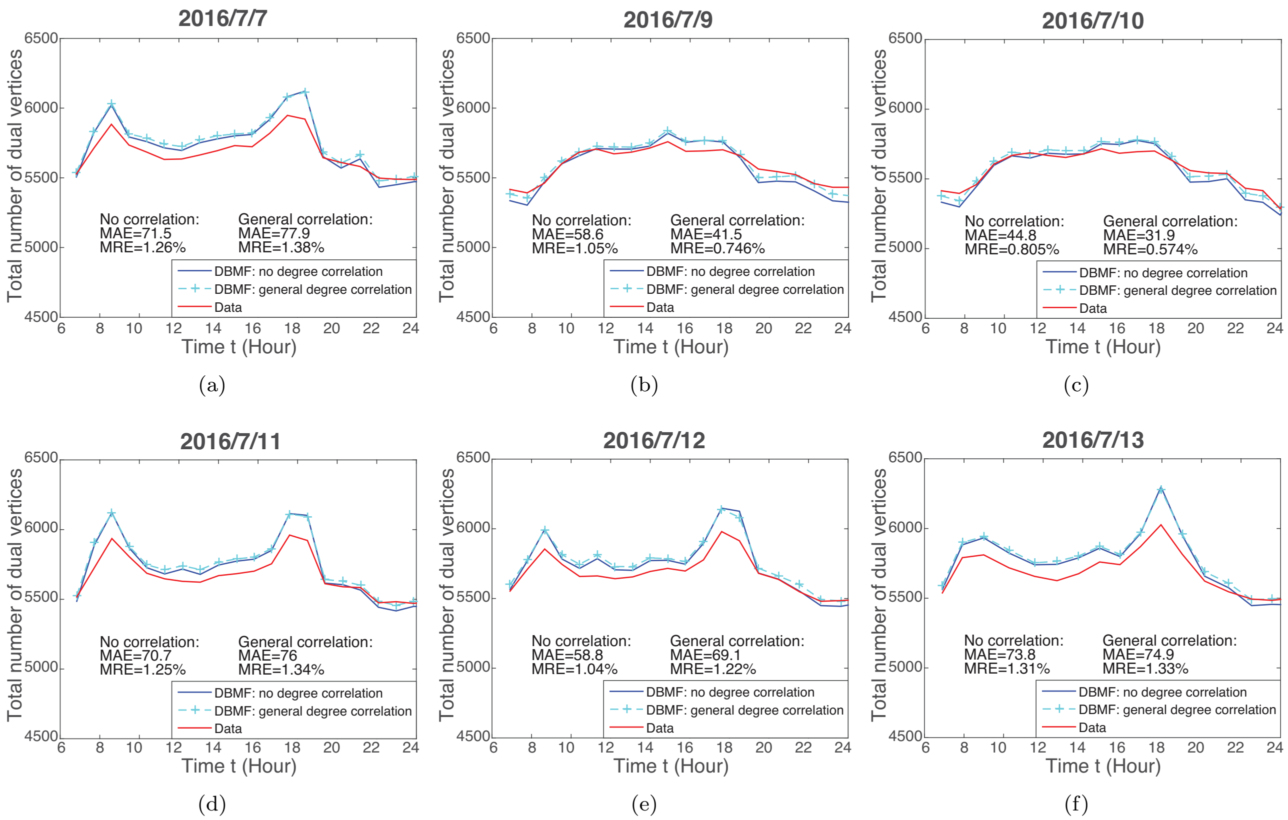

FIG. 6: Comparison of the total number of dual vertices from emprical data (red line) and the expected total number of dual vertices (N^*) obtained from the vertex split recovery model for Shanghai road network

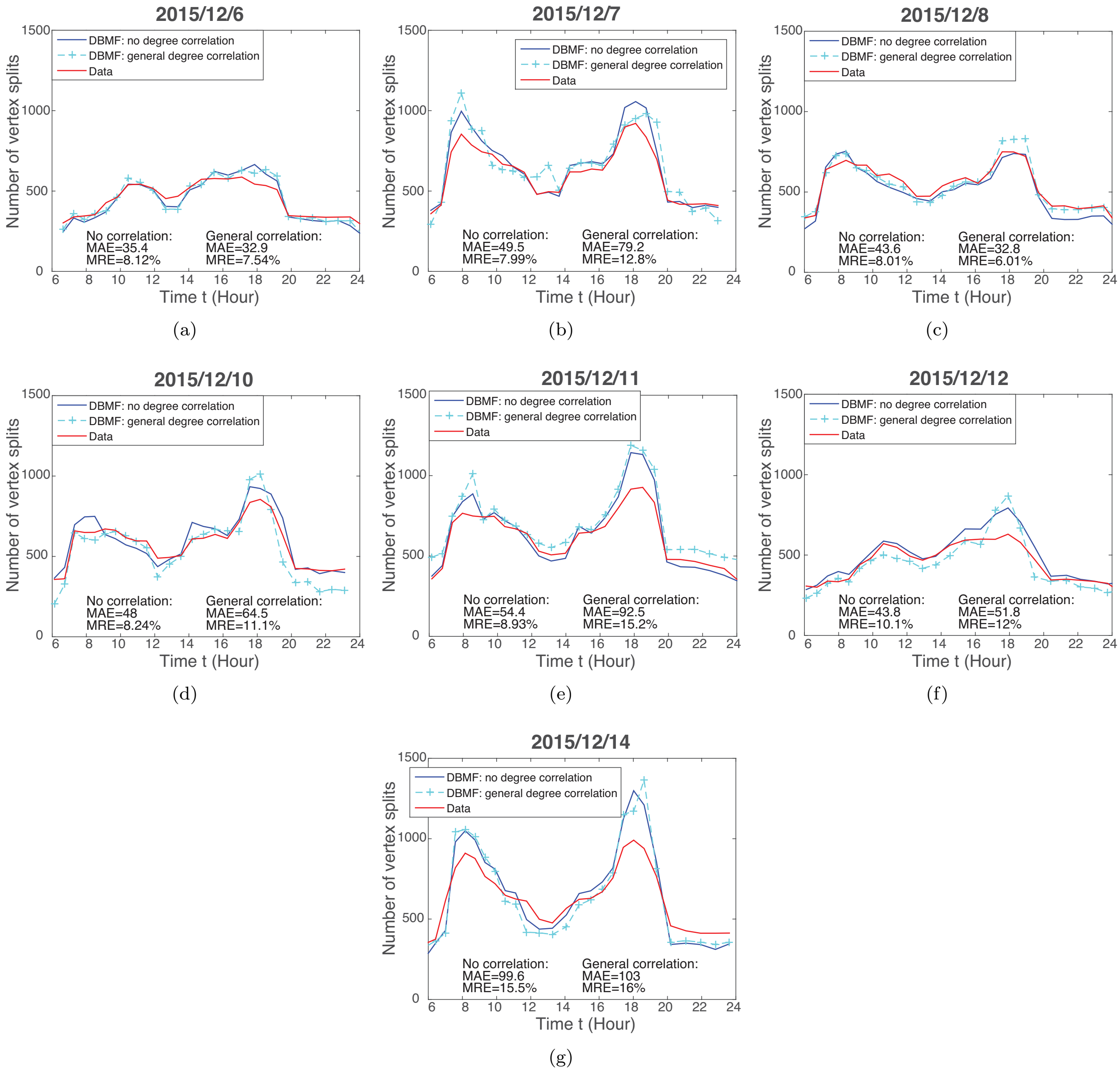

FIG. 7: Comparison of the total number of unrecovered vertex splits at each time step in empirical data (red line) and the expected number of vertex splits (N^{ES}) obtained from the vertex split recovery model for Beijing road network

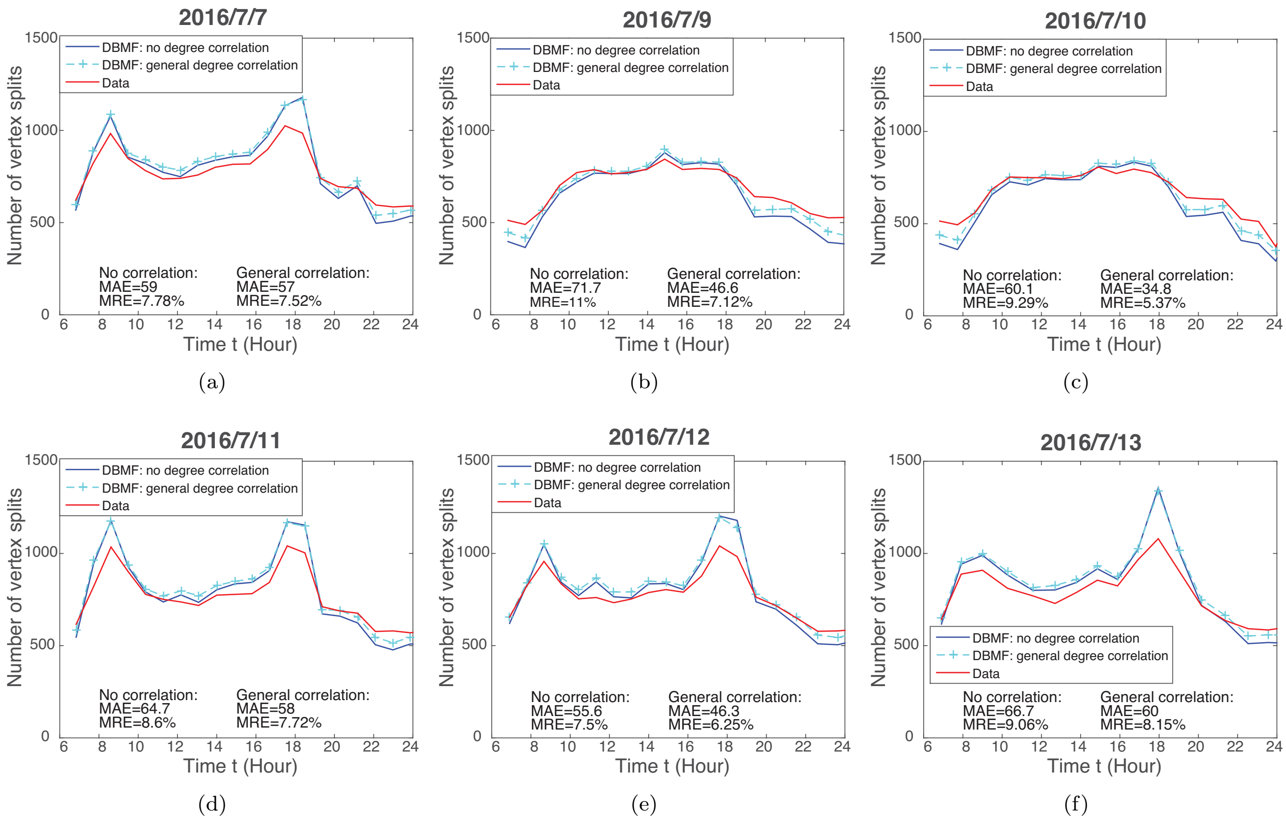

FIG. 8: Comparison of the total number of unrecovered vertex splits at each time step in empirical data (red line) and the expected number of vertex splits (N^{ES}) obtained from the vertex split recovery model for Shanghai road network

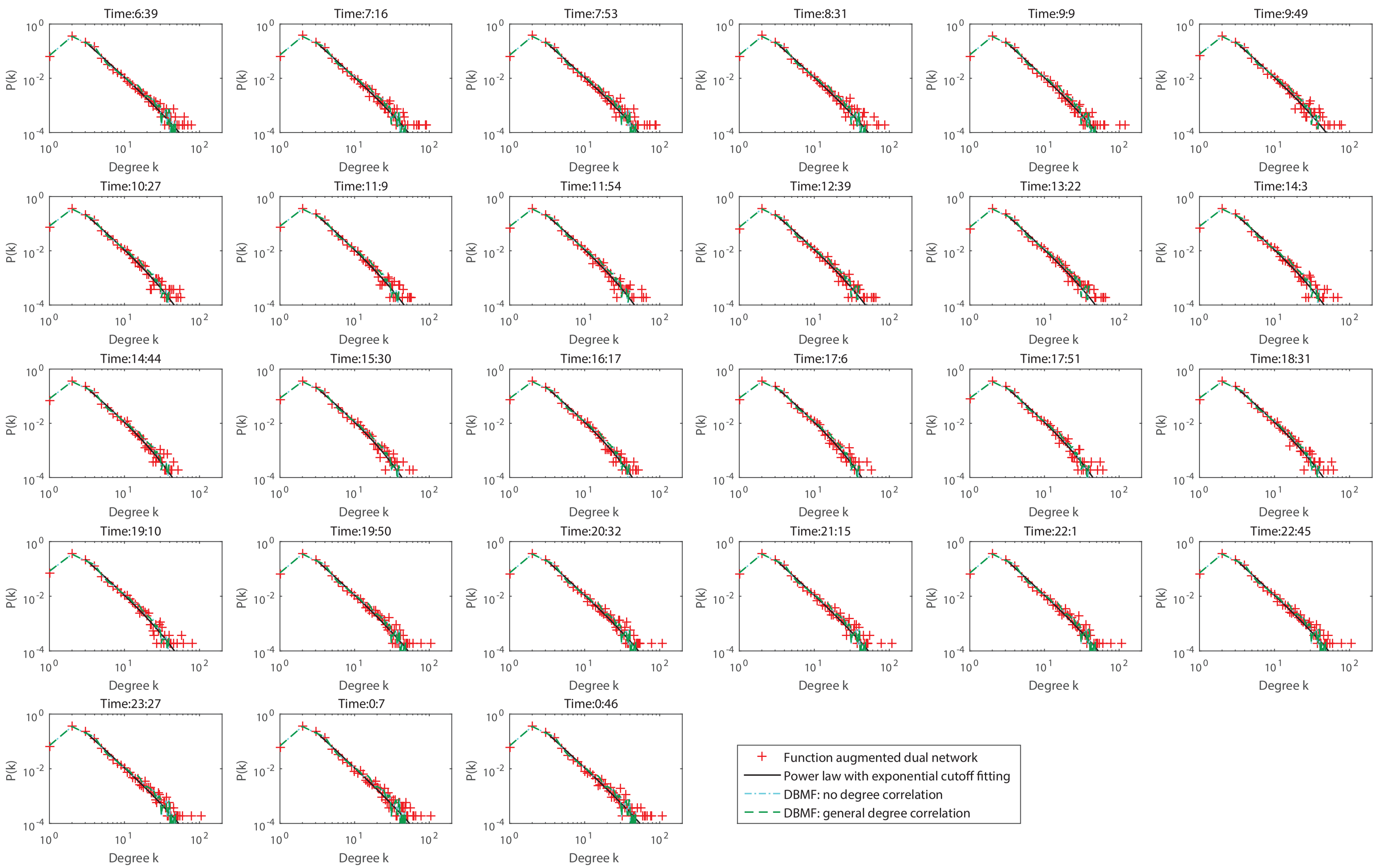

FIG. 9: Fitting results for Beijing road network: 2015/12/6

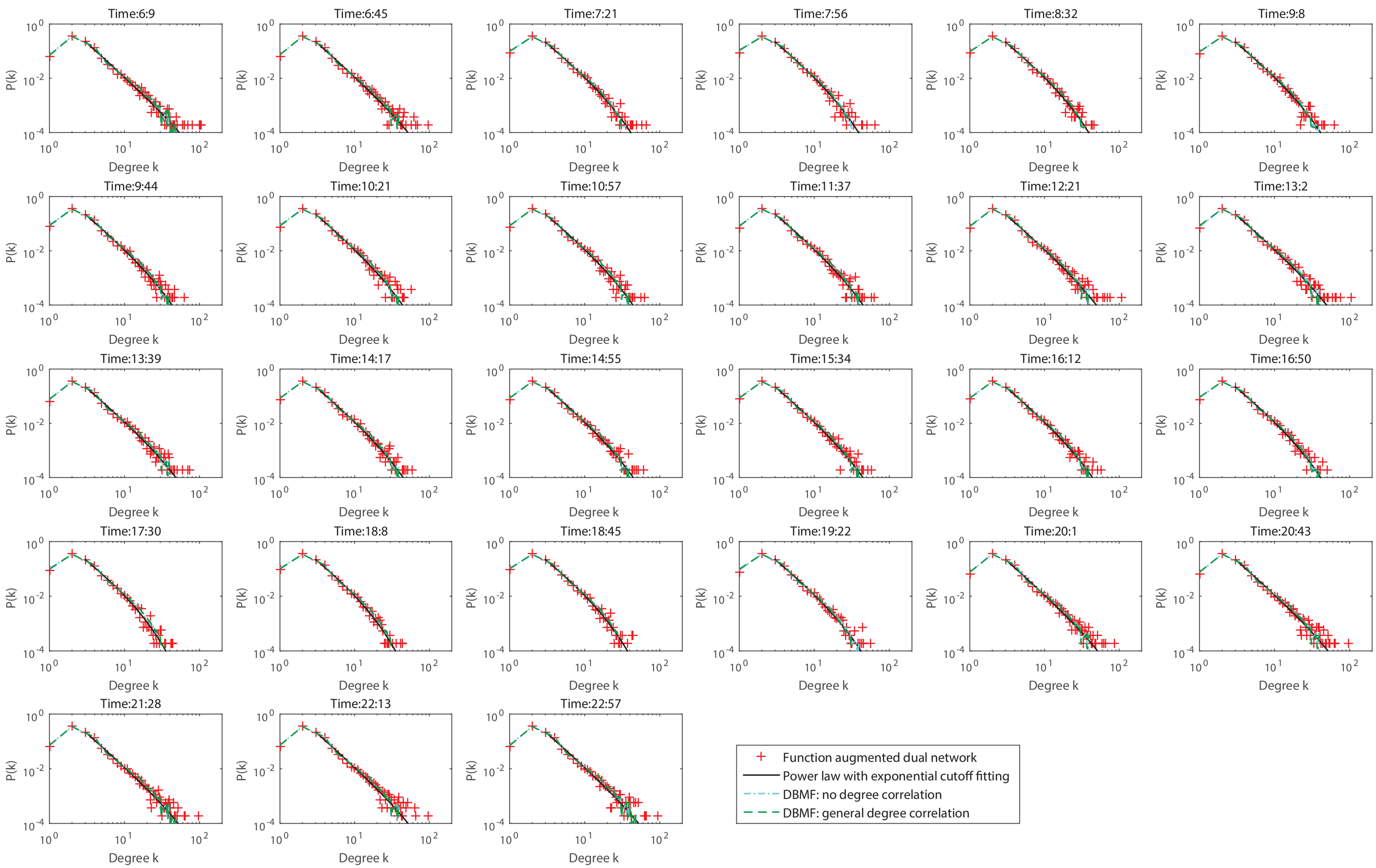

FIG. 10: Fitting results for Beijing road network: 2015/12/7

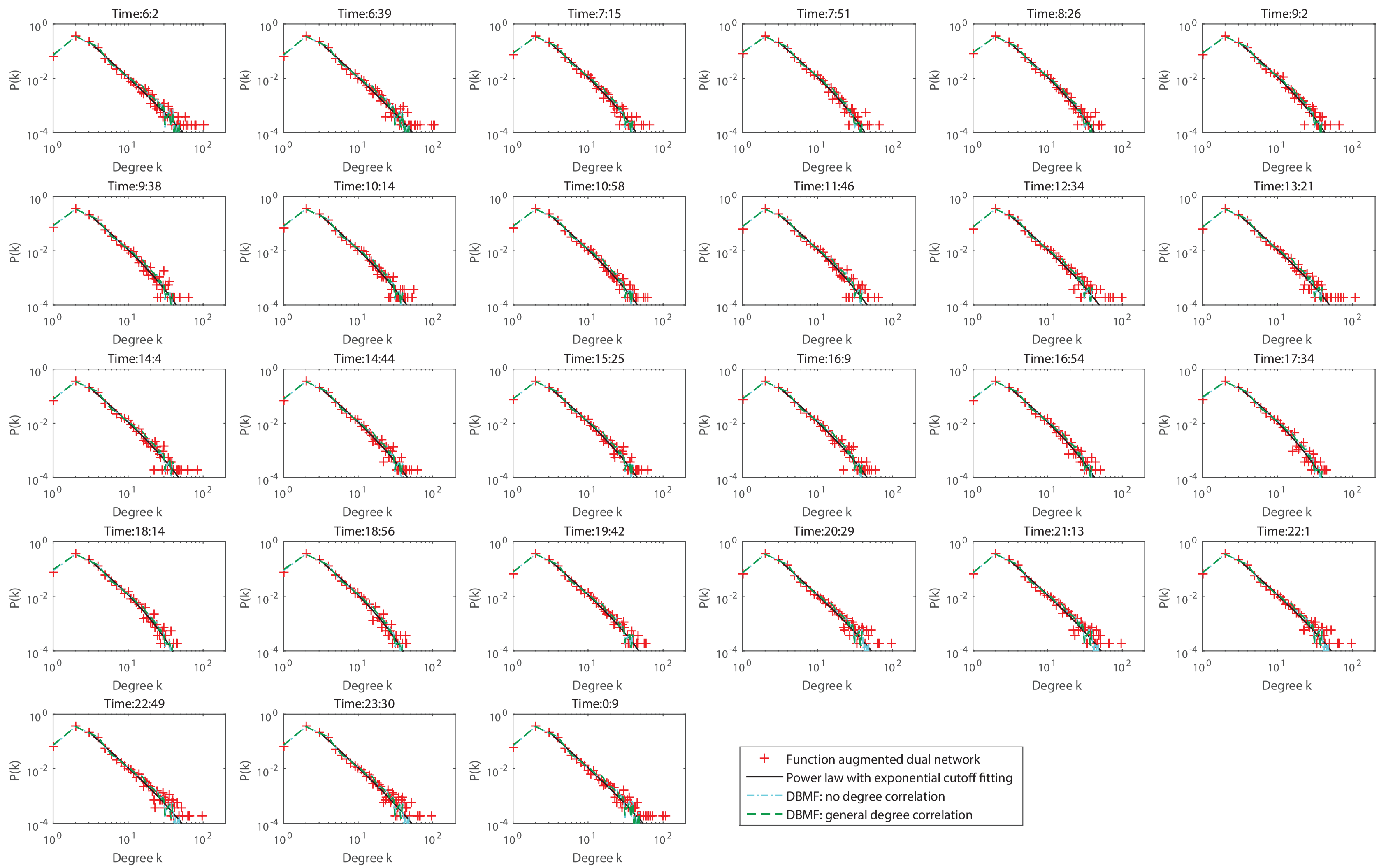

FIG. 11: Fitting results for Beijing road network: 2015/12/8

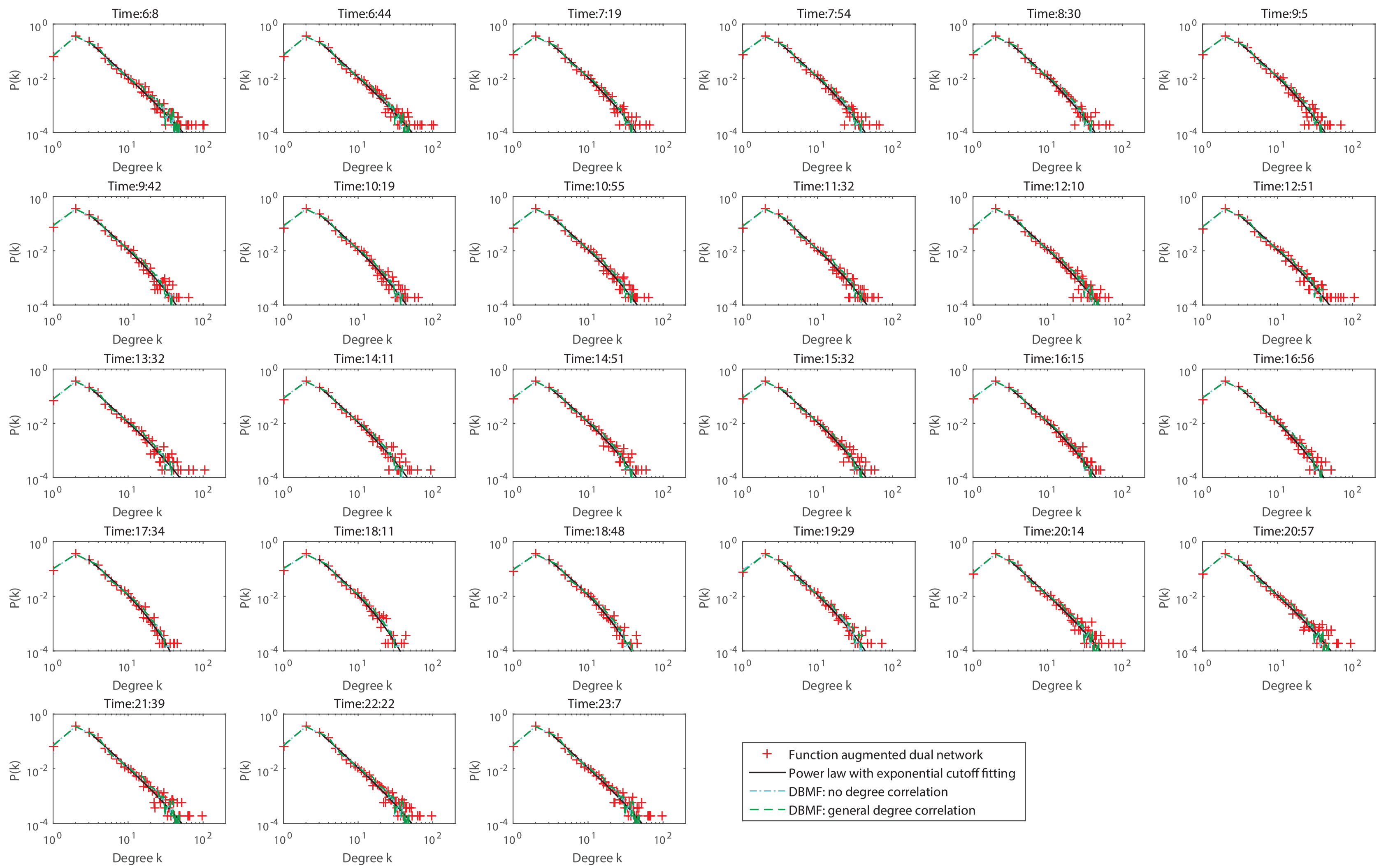

FIG. 12: Fitting results for Beijing road network: 2015/12/10

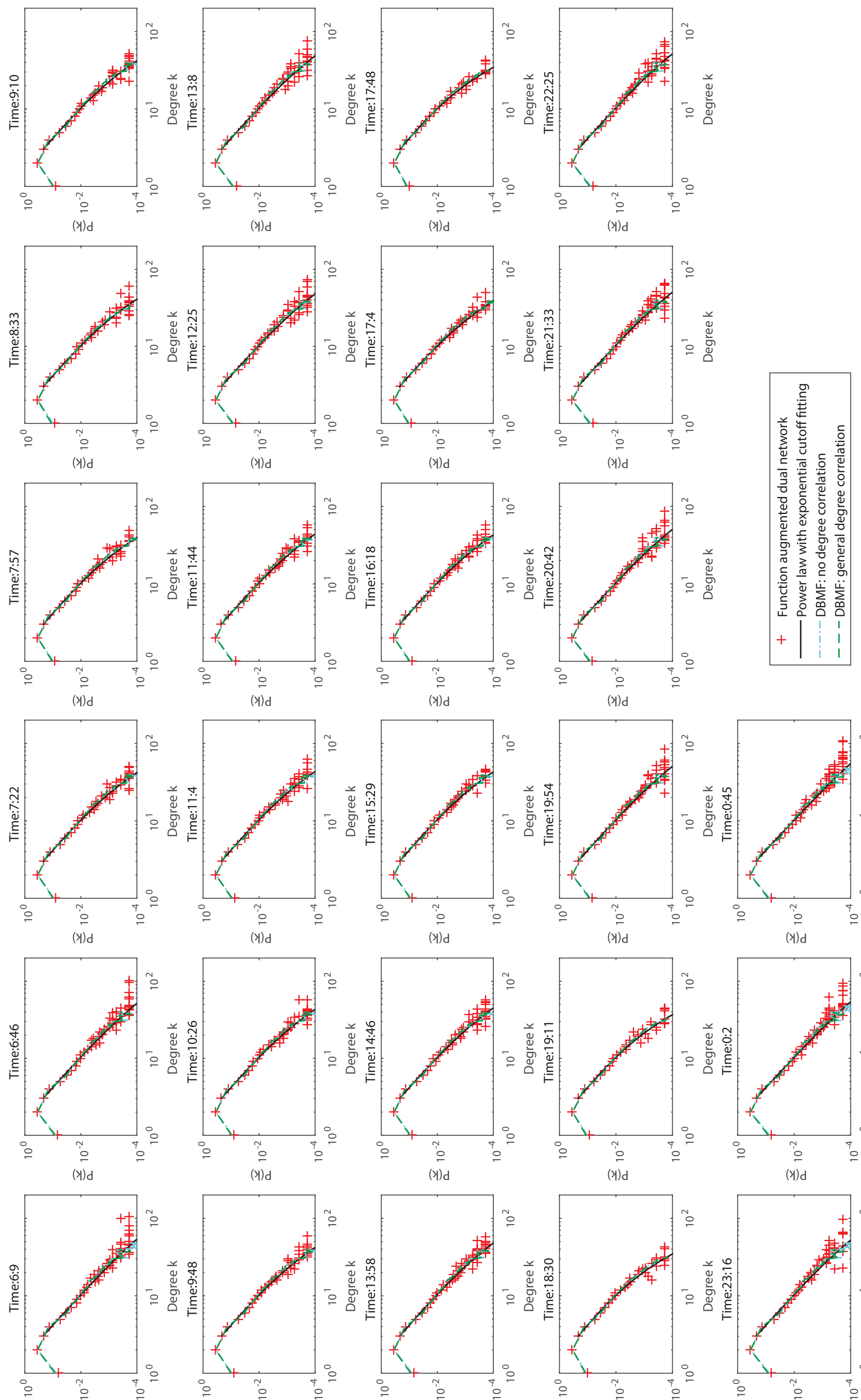

FIG. 13: Fitting results for Beijing road network: 2015/12/11

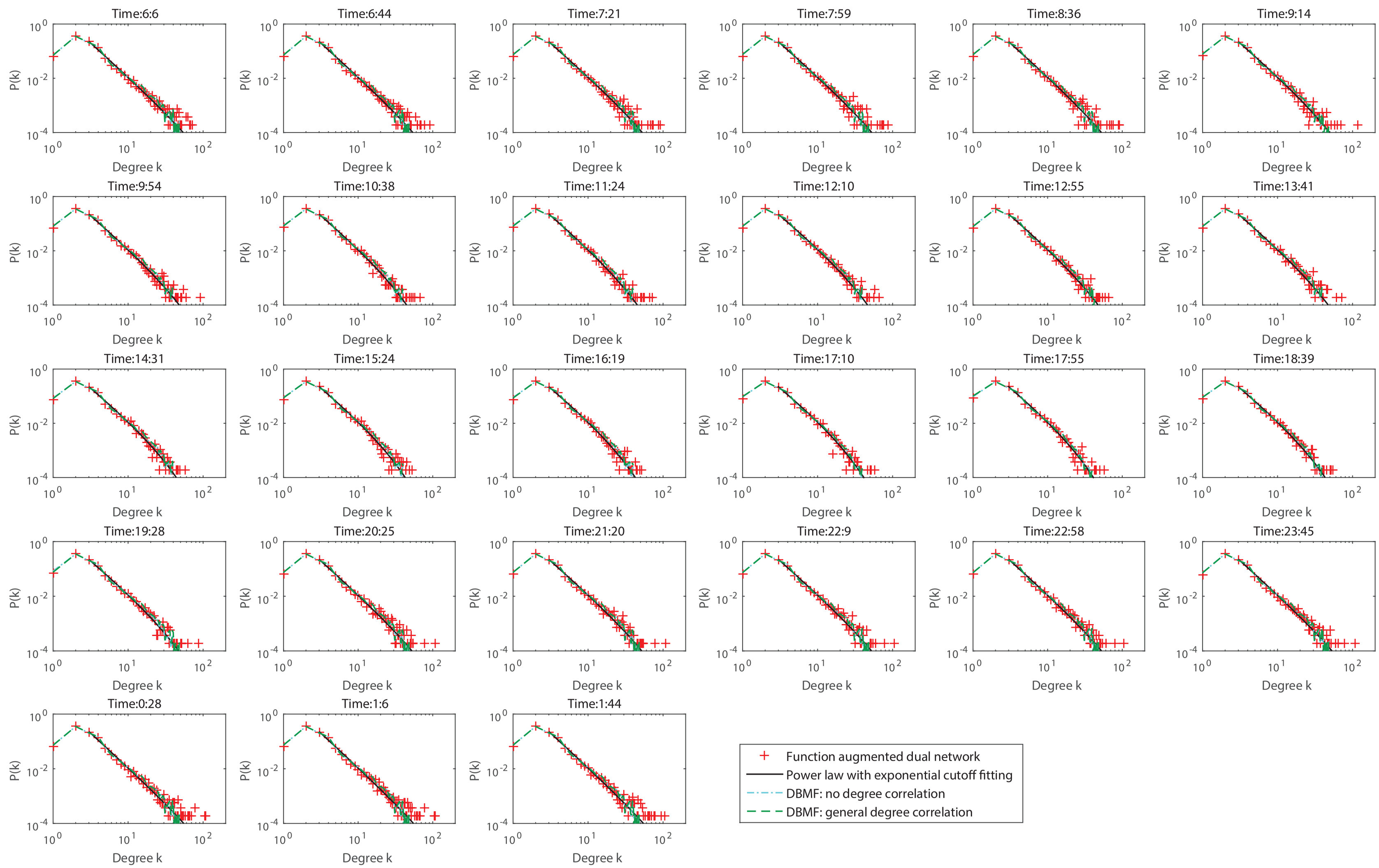

FIG. 14: Fitting results for Beijing road network: 2015/12/12

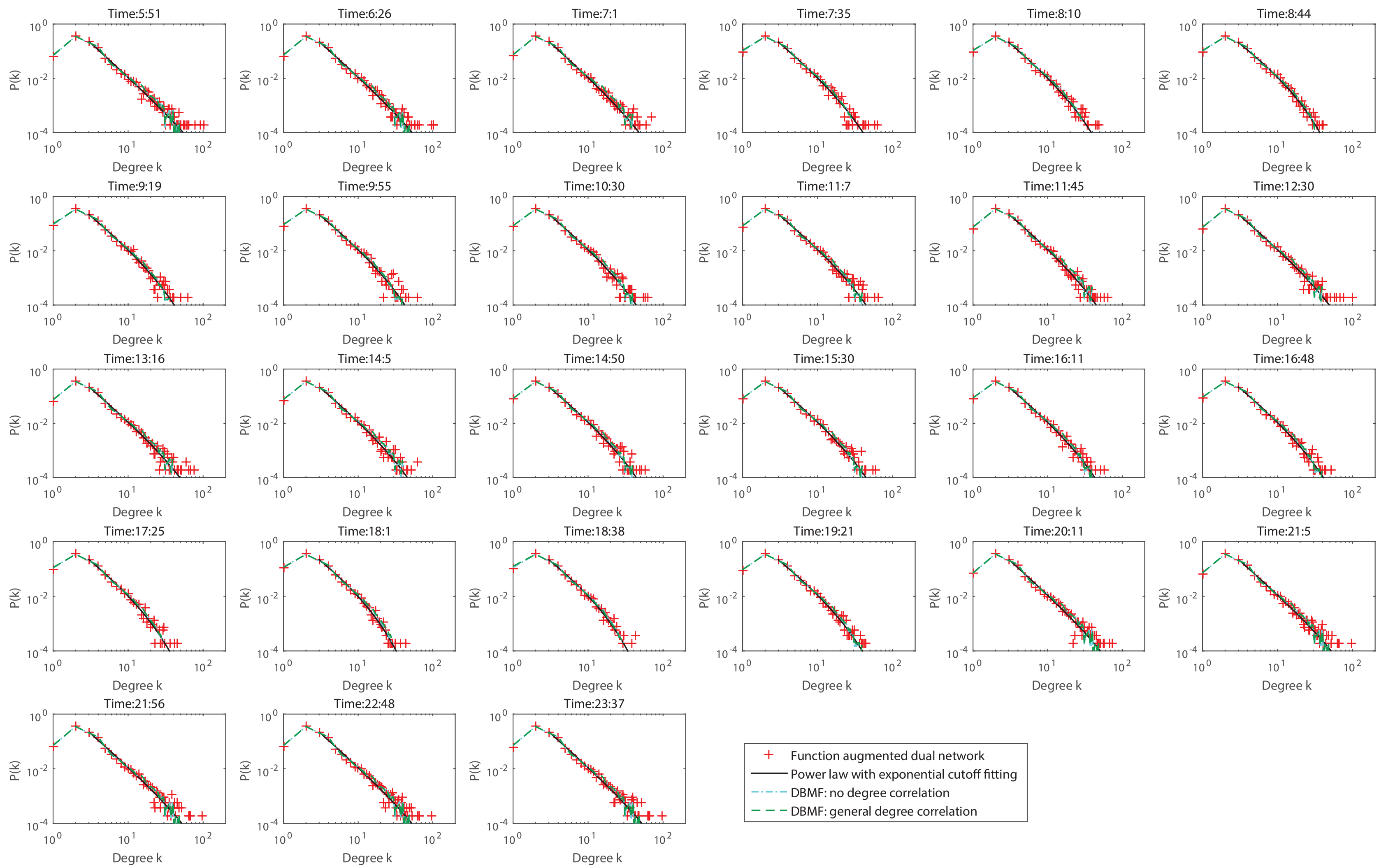

FIG. 15: Fitting results for Beijing road network: 2015/12/14

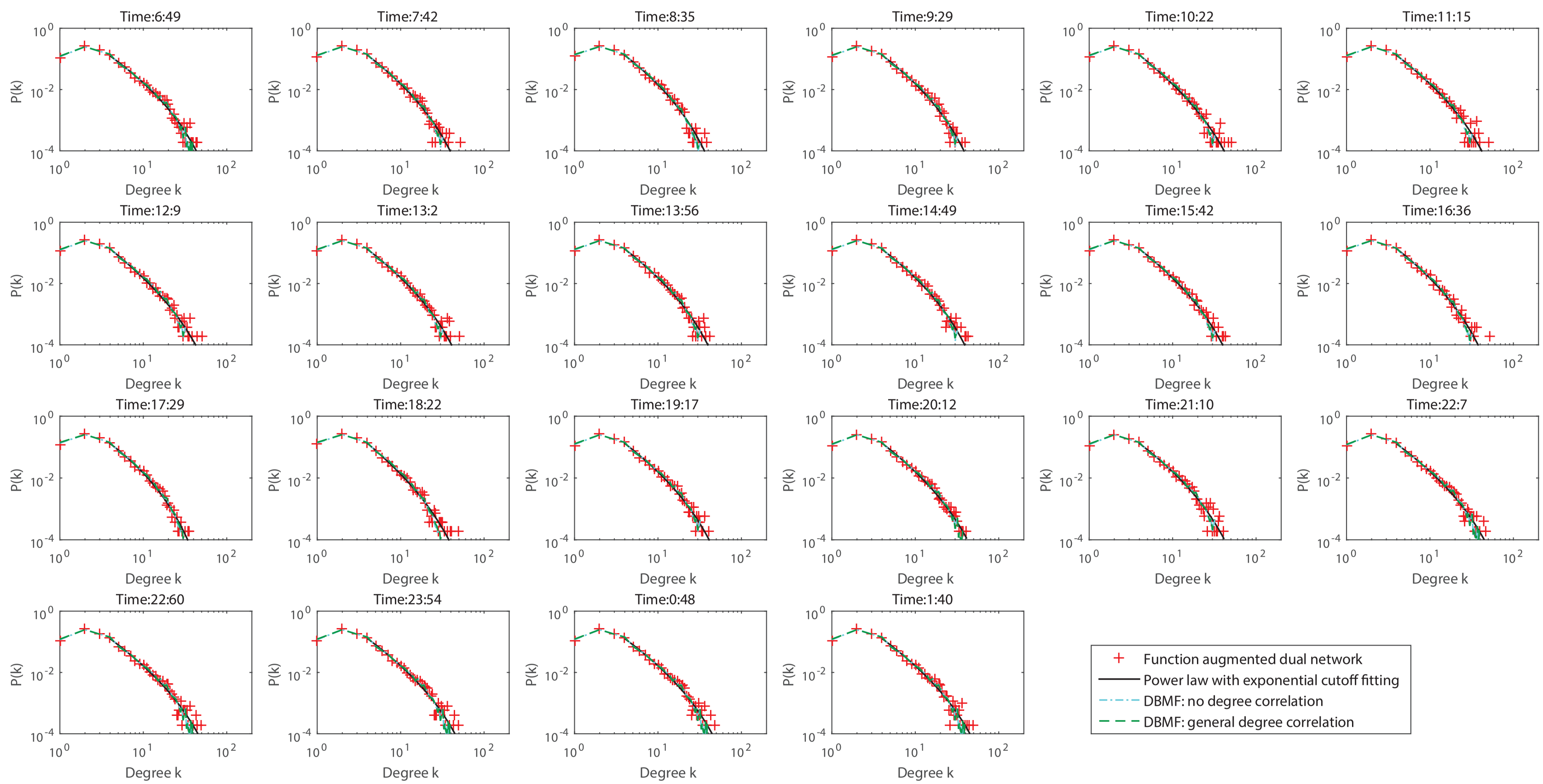

FIG. 16: Fitting results for Shanghai road network: 2016/7/7

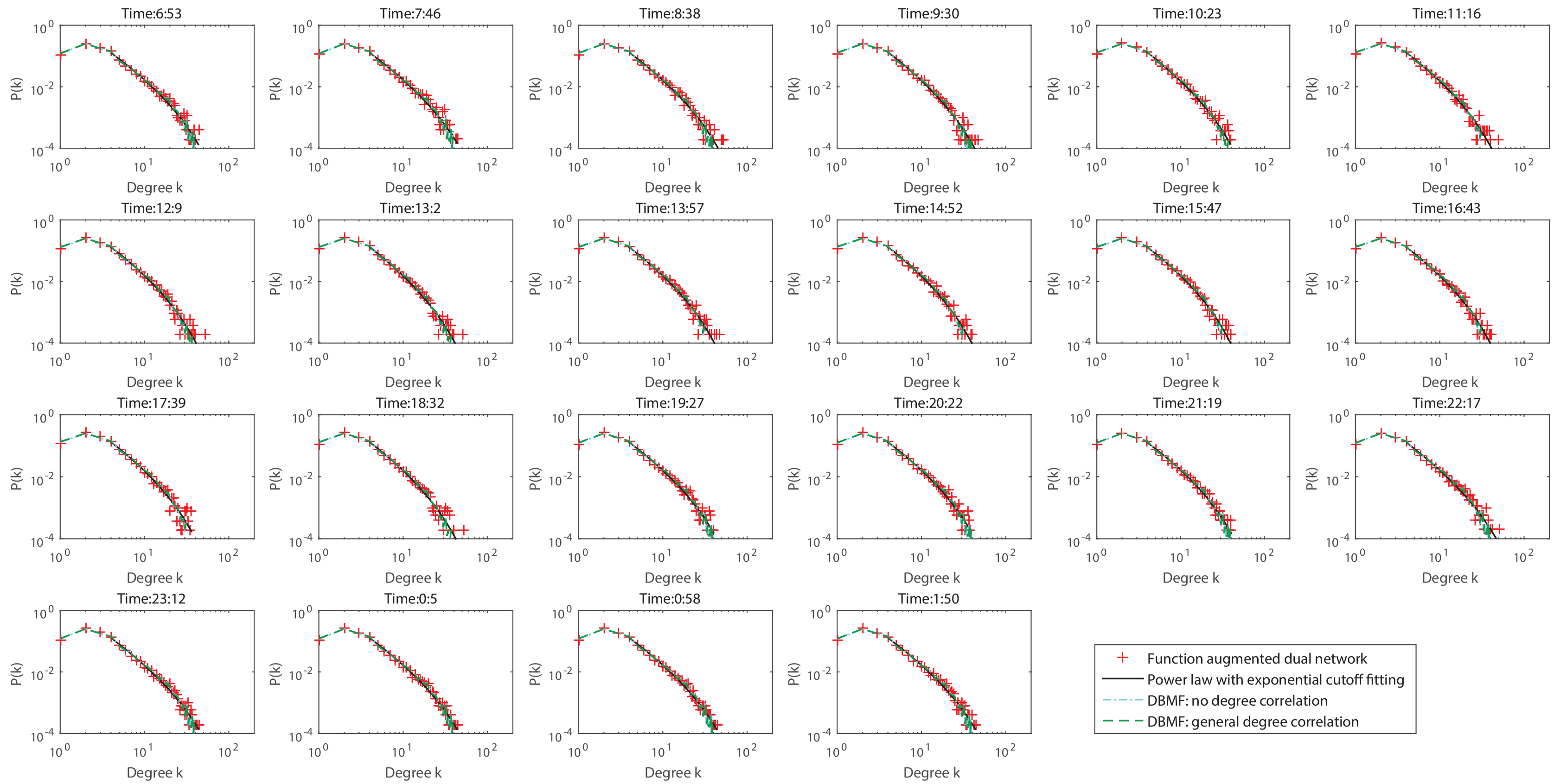

FIG. 17: Fitting results for Shanghai road network: 2016/7/9

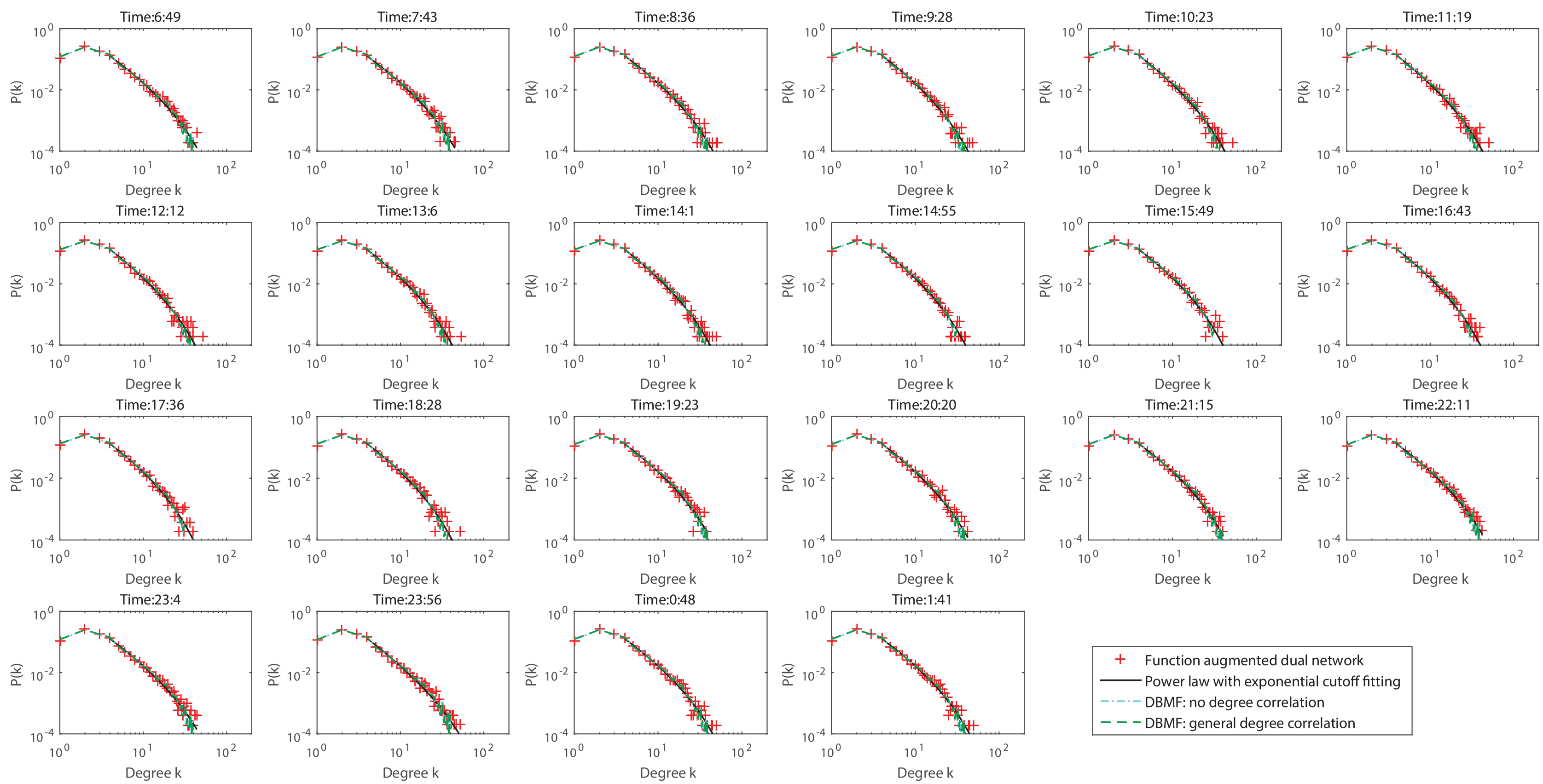

FIG. 18: Fitting results for Shanghai road network: 2016/7/10

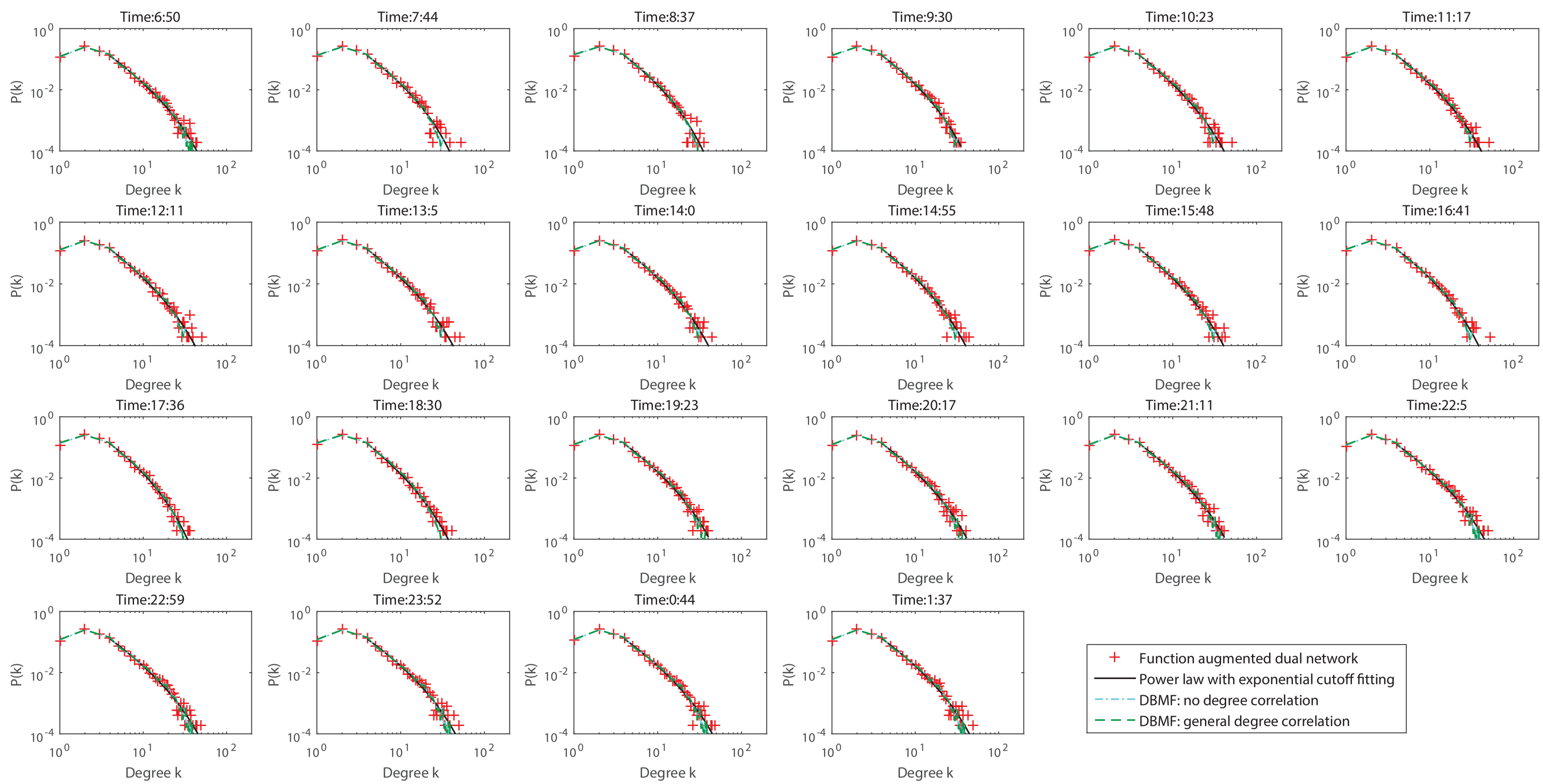

FIG. 19: Fitting results for Shanghai road network: 2016/7/11

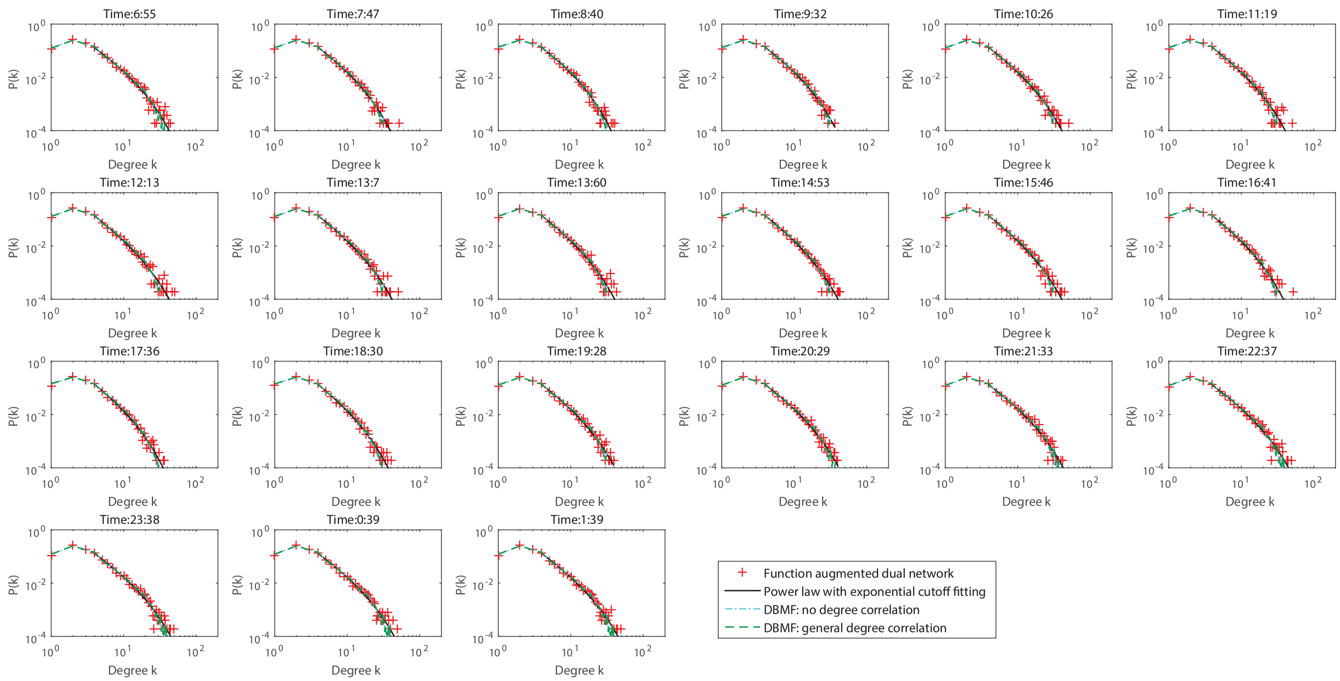

FIG. 20: Fitting results for Shanghai road network: 2016/7/12

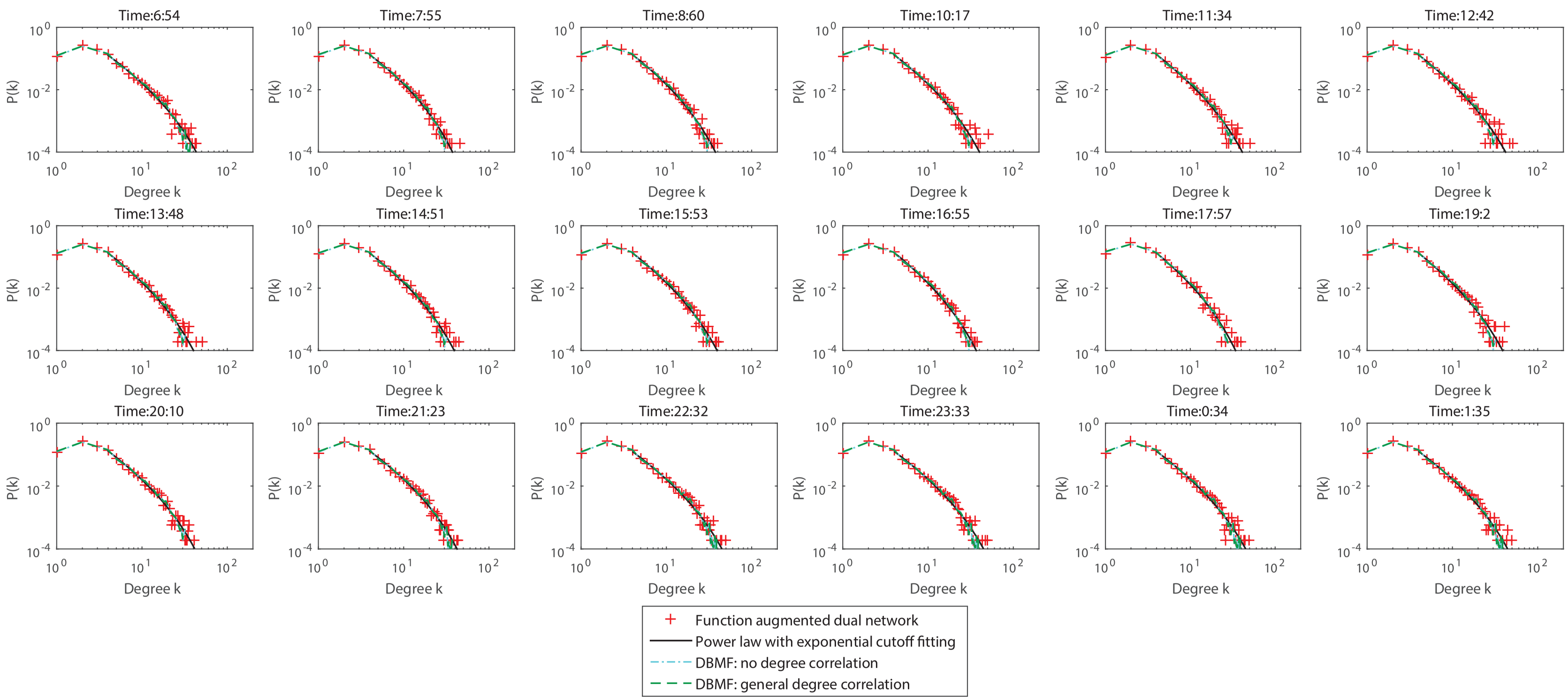

FIG. 21: Fitting results for Shanghai road network: 2016/7/13