Accuracy of mean-field theory for dynamics on real-world networks

James P. Gleeson,[1] Sergey Melnik,[1,2,3] Jonathan A. Ward,[1,4] Mason A. Porter,[2,3] and Peter J. Mucha[5,6]

[1]*MACSI, Department of Mathematics & Statistics, University of Limerick, Ireland*

[2]*Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, OX1 3LB, United Kingdom*

[3]*CABDyN Complexity Centre, University of Oxford, Oxford, OX1 1HP, United Kingdom*

[4]*Centre for Mathematics of Human Behaviour, Department of Mathematics & Statistics, University of Reading, Reading, RG6 6AX, United Kingdom*

[5]*Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, North Carolina 27599-3250, USA*

[6]*Institute for Advanced Materials, Nanoscience & Technology, University of North Carolina, Chapel Hill, North Carolina 27599-3216, USA*

Mean-field analysis is an important tool for understanding dynamics on complex networks. However, surprisingly little attention has been paid to the question of whether mean-field predictions are accurate, and this is particularly true for real-world networks with clustering and modular structure. In this paper, we compare mean-field predictions to numerical simulation results for dynamical processes running on 21 real-world networks and demonstrate that the accuracy of such theory depends not only on the mean degree of the networks but also on the mean first-neighbor degree. We show that mean-field theory can give (unexpectedly) accurate results for certain dynamics on disassortative real-world networks even when the mean degree is as low as 4.

PACS number(s): 89.75.Hc, 87.23.Ge, 89.75.Fb

I. INTRODUCTION

Mean-field theories are among the most common forms of analytical approximation employed when studying dynamics on complex networks [1]. Typically, mean-field theories are derived under several assumptions:

(i) *Absence of local clustering.* When considering possible changes to the state of a node B_1, it is assumed that the states of the neighbors of node B_1 are independent of each other. However, this assumption that the network is "locally treelike" does not hold if the neighbors of B_1 are also linked to each other—i.e., if the network is clustered (exhibits transitivity).

(ii) *Absence of modularity.* It is also usually assumed that all nodes of the same degree k are well described by the mean k-class state—i.e., by the average over all nodes of degree k. However, this might not be true if the network is modular, so that the states of degree-k nodes are differently distributed in different communities.

(iii) *Absence of dynamical correlations.* Finally, it is assumed that the states of each node B_1 and those of its neighbors can be treated as independent when updating the state of node B_1.

Importantly, the neglect of *dynamical* correlations (as distinct from *structural* correlations such as degree-degree correlations [2,3]) between neighbors in assumption (iii) can be addressed in improved theories that incorporate information on the joint distribution of node *states* at the ends of a random edge in the network [4,5] (cf. theories that only specify the *structures* at the ends of a random edge). The improved theories are often called *pair approximations* (PAs) [6,7], and these are inevitably more complicated to derive and study than mean-field (MF) theories, so we mostly restrict our attention in the present paper to the more common MF-theory situation.[1]

The distinctions between assumptions (i)–(iii) can be clarified by considering the theoretical approaches beyond the MF level that have been developed in certain cases to deal with violations of (i)–(iii). The impact of nonzero clustering on percolation problems on a network was examined in Refs. [8–12]. The analytical methods used in those papers explicitly account for the dependence of neighbors' states on each other—i.e., for the violation of MF assumption (i). The role of community or modular structures [see assumption (ii)] on percolation requires a different extension of analytical methods [13,14]. As noted above, PA theories can account for dynamical correlations more accurately than MF theory, thereby improving on assumption (iii).

The MF assumptions enumerated above are clearly violated for real-world networks, which are often highly clustered and modular [4,15]. It is therefore rather surprising that MF theory often provides a reasonably good approximation to the actual dynamics on many real-world networks. This fact has been noted by several authors [1,5,16–18], but to our knowledge no comprehensive explanation for this phenomenon has ever been developed. In studying this phenomenon, we focus on a specific question of obvious practical interest: Given a real-world network and a dynamical process running on it, is it possible to predict whether or not MF theory will provide a good approximation to the actual dynamics on this network? In this paper, we test multiple well-studied dynamical processes on 21 undirected, unweighted real-world networks. We enumerate and summarize various properties of the networks in Table I. These networks are characterized

---

[1]One should not think of PA theories as theories that only add information about degree-degree correlations. Each particular PA or

---

MF theory can either neglect degree-degree correlations (or other *structural* correlations, which specify connections between nodes) or take them into account. The difference is that MF theories are derived under assumption (iii) of the absence of *dynamical* correlations, while PA theories take the dynamical correlations into consideration.

TABLE I. Basic diagnostics and error measures for the networks used in this paper. All real-world data have been treated in the form of undirected, unweighted networks. We only consider the largest connected component of each network, for which we enumerate the following properties: total number of nodes N; mean degree z; mean first-neighbor degree d; fraction of nodes of degree 10 or less; clustering coefficients C and $\widetilde{C}$ (defined, respectively, in Eqs. (3.6) and (3.4) of Ref. [48]); and Pearson degree correlation coefficient r. The quantities E_S, E_K, and E_V are the relative MF errors for the SIS, Kuramoto, and voter models; and E_V^{PA} is the relative PA error for the voter model. The last column gives the citation for the network in this paper's bibliography and/or the URL of a data file.

No.	Network	N	z	d	frac($k \leqslant 10$)	C	$\widetilde{C}$	r	E_K	E_S	E_V	E_V^{PA}	Refs.
1	Word adjacency: Spanish	11558	7.45	942.58	0.91	0.38	0.02	-0.2819	-0.03	0.01	0.21	0.04	[25]
2	Word adjacency: English	7377	11.98	666.03	0.81	0.41	0.04	-0.2366	-0.03	0.01	0.12	0.03	[25]
3	AS Internet	28311	4.00	473.65	0.96	0.21	0.01	-0.2000	0.08	0.03	0.44	0.08	[26]
4	Word adjacency: French	8308	5.74	376.22	0.93	0.21	0.01	-0.2330	-0.02	0.03	0.23	0.003	[25]
5	Marvel comics	6449	52.17	338.16	0.25	0.78	0.19	-0.1647	-0.03	0.02	0.03	0.008	[27]
6	Reuters 9/11 news	13308	22.25	236.17	0.70	0.37	0.11	-0.1090	-0.01	0.01	0.04	-0.009	[28]
7	Word adjacency: Japanese	2698	5.93	199.99	0.92	0.22	0.03	-0.2590	-0.02	0.02	0.24	0.01	[25]
8	Facebook Oklahoma	17420	102.47	186.04	0.11	0.23	0.16	0.0737	0.03	0.01	0.02	0.01	[19,29]
9	Corporate ownership (EVA)	4475	2.08	113.85	0.98	0.01	0.00	-0.1851	0.80	0.06	1.00	0.65	[30]
10	Political blogs	1222	27.36	100.07	0.45	0.32	0.23	-0.2213	0.04	0.01	0.05	0.009	[31]
11	Facebook Caltech	762	43.70	74.65	0.20	0.41	0.29	-0.0662	0.004	0.01	0.04	0.009	[19,29]
12	Airports500	500	11.92	59.50	0.76	0.62	0.35	-0.2679	-0.002	0.06	0.29	0.19	[32,33]
13	*C. Elegans* metabolic	453	8.94	51.57	0.86	0.65	0.12	-0.2258	0.03	0.04	0.18	0.04	[34,35]
14	Interacting proteins	4713	6.30	32.92	0.84	0.09	0.06	-0.1360	0.14	0.04	0.19	-0.03	[36–38]
15	*C. Elegans* neural	297	14.46	32.00	0.41	0.29	0.18	-0.1632	0.10	0.03	0.09	0.006	[20,39]
16	Transcription yeast	662	3.21	22.31	0.95	0.05	0.02	-0.4098	0.64	0.08	0.70	0.36	[40]
17	Coauthorships	39577	8.88	20.17	0.77	0.65	0.25	0.1863	0.22	0.08	0.14	0.01	[41,42]
18	Transcription *E. coli*	328	2.78	17.88	0.96	0.11	0.02	-0.2648	0.48	0.09	0.79	0.43	[43]
19	PGP Network	10680	4.55	13.46	0.90	0.27	0.38	0.2382	0.50	0.16	0.45	0.17	[44–46]
20	Electronic Circuit (s838)	512	3.20	4.01	1.00	0.05	0.05	-0.0300	0.78	0.23	0.62	0.23	[25]
21	Power Grid	4941	2.67	3.97	0.99	0.08	0.10	0.0035	0.93	0.29	0.90	0.65	[20,47]
22	γ theory [$\gamma(3,3)=1$]	1002	3	3	1.00	1/3	1/3	N/A	0.91	0.89	0.78	0.47	[10]
23	γ theory [$\gamma(3,3)=1$]	10002	3	3	1.00	1/3	1/3	N/A	0.97	0.88	0.79	0.49	[10]

by a range of values for several standard network diagnostics, which is important for this study. We show that MF theory typically works well provided that d, the mean degree of first neighbors of a random node, is sufficiently large. In contrast, we demonstrate that the mean degree z of a network is not necessarily a good indicator of MF accuracy.

The remainder of this paper is organized as follows. In Sec. II, we introduce the dynamical processes that we consider and compare numerical results with MF theory for sample real-world networks. In Sec. III, we discuss the implications of our results and propose an explanatory hypothesis. In Sec. IV, we compare our results with earlier work in this area. We conclude in Sec. V.

II. EXAMPLES

We begin by showing examples for which MF theory gives accurate results for dynamics on real-world networks and contrast them with examples for which MF theory performs poorly.

A. Kuramoto phase oscillator model

In Fig. 1, we show the results of running the Kuramoto phase oscillator model [22] on the Facebook Oklahoma network [19] and on the US Power Grid network [20]. Each node corresponds to an oscillator with an intrinsic frequency drawn

from a unit-variance Gaussian distribution. The phase $\theta_j(t)$ of the oscillator at node j obeys the differential equation

$$\frac{d\theta_j}{dt} = \omega_j + K \sum_{\ell=1}^{N} A_{j\ell} \sin(\theta_\ell - \theta_j), \qquad (1)$$

where ω_j is the intrinsic frequency of node j, N is the number of nodes, and A is the adjacency matrix of the network. The
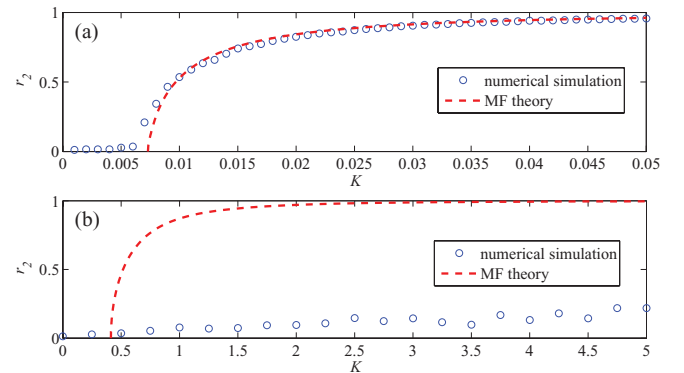

FIG. 1. (Color online) Order parameter for synchronization in the Kuramoto phase oscillator model running on (a) the Facebook Oklahoma network [19] and (b) the US Power Grid network [20] as a function of the coupling parameter K. The order parameter r_2 is defined in Eq. (2). The MF theory of [21] is given by Eq. (3).

coupling to network neighbors is measured by the parameter K, and global synchrony of the oscillators is expected to emerge for sufficiently large K [23]. Synchrony is quantified using the order parameter [21]

$$r_2 = \frac{\left| \sum_{j=1}^N k_j e^{i\theta_j} \right|}{\sum_{j=1}^N k_j},$$ (2)

where k_j is the degree of node j and $i = \sqrt{-1}$. The MF theory of Ref. [21] (see also [24]) yields the following implicit equation for r_2:

$$\sum_{k=0}^{\infty} k^2 p_k e^{-K^2 k^2 r_2^2/4} \left[I_0\left(\frac{K^2 k^2 r_2^2}{4}\right) + I_1\left(\frac{K^2 k^2 r_2^2}{4}\right) \right] = \frac{2\sqrt{2}z}{\sqrt{\pi}K},$$ (3)

where I_n is the modified Bessel function of the first kind, p_k is the degree distribution of the network, and $z = \langle k \rangle \equiv \sum k p_k$ is the mean degree. The agreement in Fig. 1 between theory and simulation is very good for the Facebook network but very poor for the Power Grid. (See Fig. 4 for additional examples.)

The results of Fig. 1 are perhaps explained in part by noting that the mean degree z of the Facebook network is about 102, whereas $z \approx 2.67$ for the Power Grid (see Table I). It is arguable that nodes with many neighbors will experience something closer to a "mean field" than nodes with few neighbors. In particular, it is plausible that low-z networks might be more prone to errors in MF due to neglecting the effects of clustering, modularity, and dynamical correlations. This is attractively simple, but as we show below, this naive explanation does not fully capture certain subtleties.

B. SIS epidemic model

In Fig. 2, we compare simulations for the susceptible-infected-susceptible (SIS) epidemic model [51–53] with the corresponding predictions of a well-known MF theory [49,50]. In this MF theory, the fraction $i_k(t)$ of degree-k nodes that are infected at time t is given by the solution of the equation

$$\frac{di_k}{dt} = -i_k + \beta(1 - i_k)k\Theta_k,$$ (4)

where β is the spreading rate, the recovery rate has been set to unity by choice of time scale, and

$$\Theta_k = \sum_{k'} P(k'|k) i_{k'}$$ (5)

is the probability that any given neighbor of a degree-k node is infected. In Eq. (5), $P(k'|k)$ is the probability that an edge originating at a degree-k node has a degree-k' node at its other end. Because degree-degree correlations are included, this version of the theory is called a *correlated MF theory* (cMF). A further simplification of the theory is possible if one assumes that the network is uncorrelated and is thus completely described by its degree distribution p_k. In this case, which is termed *uncorrelated MF* (uMF), $P(k'|k)$ in Eq. (5) is replaced by $k'p_{k'}/z$ and Θ_k becomes independent of k. In
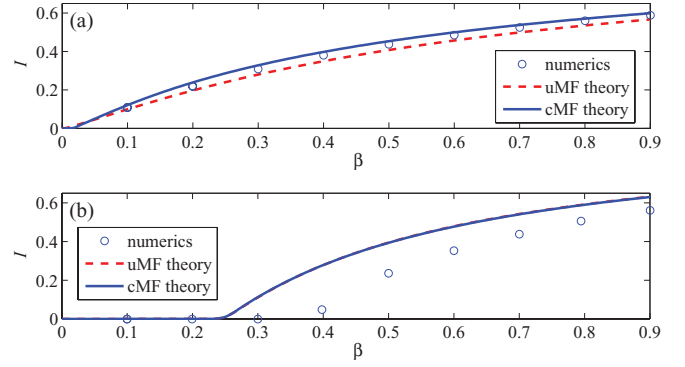

FIG. 2. (Color online) Fraction of infected nodes in the steady state of the SIS process on the (a) Autonomous System (AS) Internet [26] and (b) Electronic Circuit (s838) [25] networks as a function of the spreading rate β. The MF theory is from Refs. [49,50]. Observe that uncorrelated and correlated MF theories are indistinguishable in panel (b).

Fig. 2, we show predictions of both correlated and uncorrelated MF theories for the steady-state endemic infected fraction

$$I = \lim_{t \to \infty} \sum_k p_k i_k(t)$$

for the AS Internet network [26] and for the Electronic Circuit (s838) network [25]. The MF theory is very accurate for the former, but it performs poorly for the latter.

The result for the AS Internet network is particularly surprising in light of the fact that the mean degree of this network is only 4. The aforementioned naive argument that MF theory is accurate for high-degree nodes thus cannot account for the good performance of the theory in this low-z case, in which 96% of the nodes in the network have degree 10 or less (see Table I).

C. Voter model

As a third example, we consider the survival probability of disordered-state trajectories in the voter model [54] and compare it with the (uncorrelated) MF theory of Ref. [55]. (For rigorous results for the voter model, see Refs. [5,56] and references therein.) At time $t = 0$, each node is randomly assigned to one of two voter states. In each time step (of size $dt = 1/N$), a randomly chosen node is updated by copying the state of one of its randomly chosen neighbors. On finite networks, the dynamics eventually drive a connected component to a situation of complete order, in which all of the nodes are in the same state. The survival probability $P_s(t)$ is defined as the fraction of realizations that remain in the disordered state at time t. The red dashed curves in Fig. 3 give the survival probability predicted by the MF theory of Ref. [55]:

$$P_s(t) \sim \frac{3}{2} \exp\left(-\frac{2\langle k^2 \rangle}{z^2} \frac{t}{N}\right) \quad \text{as} \quad t \to \infty,$$ (6)

and the black solid curve gives the results of the pair approximation (PA) theory of Ref. [6]:

$$P_s(t) \sim \frac{3}{2} \exp\left(-\frac{2(z-2)\langle k^2 \rangle}{(z-1)z^2} \frac{t}{N}\right) \quad \text{as} \quad t \to \infty.$$ (7)
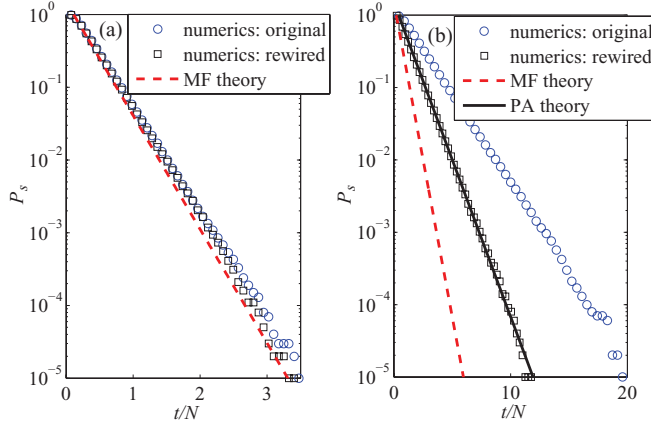
FIG. 3. (Color online) Survival probability for voter-model trajectories in the disordered state as a function of time on (a) *C. elegans* neural network [20] and (b) a synthetic clustered network with $\gamma(3,3) = 1$, which is generated as described briefly in the main text (see Ref. [10] for details). Both networks contain $N \approx 300$ nodes. The theory curves are from Refs. [6,55].

In Fig. 3(a), we show results for the *C. elegans* neural network [20] ($z \approx 14.46$), for which MF theory is very accurate. In Fig. 3(b), we show results for a synthetic clustered network described in Ref. [10]: Every node in the network has degree 3 and is part of a single 3-clique. Using the notation of Ref. [10], this is called a $\gamma(3,3) = 1$ network. It can alternatively be described in the notation of Ref. [8] as having $p_{1,1} = 1$, as every node is part of one triangle and has one single edge other than those belonging to the triangle. Clustering has a strong effect in this (low-z) network; because of this clustering and dynamical correlations, MF theory is very inaccurate. We can make the clustering negligibly small while keeping the degree distribution unchanged (i.e., $p_k = \delta_{k,3}$) by randomly rewiring the network to give a random 3-regular graph. However, even after this rewiring, the match to MF theory is poor because dynamical correlations are still neglected. The recent PA theory for the voter model [6] (see also [7]) accounts for the dynamical correlations and hence gives a good match to the survival probability on the rewired network but not on the original clustered network.

III. WHY IS MEAN-FIELD THEORY ACCURATE?

Briefly summarizing our observations thus far, we have seen (i) situations in which high-z networks exhibit good matches to MF theory but also (ii) some examples in which low-z networks have accurate MF theories. We have discussed only a few examples in detail, but these are representative of behavior observed for different dynamical processes on a variety of real-world networks. In Fig. 4, we show additional examples for each of the three dynamical processes (the Kuramoto, SIS, and voter models) for each of the six networks used in Figs. 1–3.

A. Mean-field accuracy for the SIS model

Clearly, the success of MF theories for dynamical systems on networks cannot be explained purely in terms of the mean

degree z of the underlying network. Figure 2(a) in particular gives an example in which MF theory works well on a low-z network. To understand this seemingly surprising accuracy, we focus on the SIS model and consider how the state of a node is updated as compared to the assumptions of MF theory. Suppose the state of a degree-k node B_1 is being updated. In both the true dynamics and in MF theory, the updating process depends on the state of the neighbors of B_1. Let us take node B_2 as a representative neighbor of B_1 and suppose that B_2 has degree k'. Under MF assumption (iii), the state of node B_2 is considered to be independent of the state of node B_1. This is why Eq. (5) involves the term $i_{k'}$, which is the probability that degree-k' nodes are in the infected state, without any conditioning on the state of their neighbor B_1.[2]

In reality, however, the states of nodes B_1 and B_2 exhibit dynamical correlations. For example, during an epidemic, an infected node is more likely to have infected neighbors than a susceptible node. Such dynamical correlations can be included explicitly in pair-approximation theories [57–59], and their neglect can be a major source of error in MF theories. This suggests an important question: Under what circumstances might the MF assumption of dynamical independence (iii) still give accurate results for the update of node B_1? One can argue that if the degree k' of node B_2 is sufficiently large, then the state of node B_2 is influenced by many of its neighbors other than node B_1, so the error in neglecting the particular dynamical correlation between B_2 and B_1 is relatively small for the purpose of updating node B_1. Conversely, if the degree k' of node B_2 is small, then node B_1 has a relatively strong influence on the state of node B_2, and neglecting dynamical correlations between B_2 and B_1 when updating B_1 will yield large errors. Hence, we expect MF theory to give reasonably accurate updates for node B_1 if its neighbors have sufficiently high degrees. Importantly, this argument relies only on the degree k' of the nearest-neighbor nodes being high and gives no restriction on the degree k of the updating node itself. (We remark that the use of networks with nearest-neighbor nodes of high degree has also been mentioned in studies of the Kuramoto model on networks [23,24].)

In short, we argue that MF theory gives relatively small errors for nodes with high-degree neighbors. More specifically, MF theory is likely to be *inaccurate* if many nodes do not have high-degree neighbors. This can happen, for example, if low-degree nodes are connected preferentially to other low-degree nodes (a sort of "poor-club phenomenon," akin to the rich-club phenomenon of high-degree nodes connecting preferentially to other high-degree nodes [37]). This suggests a simple but effective predictor of MF accuracy: If the mean degree of first neighbors $d = \sum_k p_k \sum_{k'} k' P(k'|k)$ is high, then MF theory can be expected to be accurate. As we show below, this rule of thumb works well for SIS dynamics on all of the networks that we have considered.

---

[2]Assumption (i) also appears in Eq. (5) through the summation of $i_{k'}$. This supposes that the neighbors' states are independent of each other—i.e., that pairs of neighbors do not form part of a triangle with node B_1.
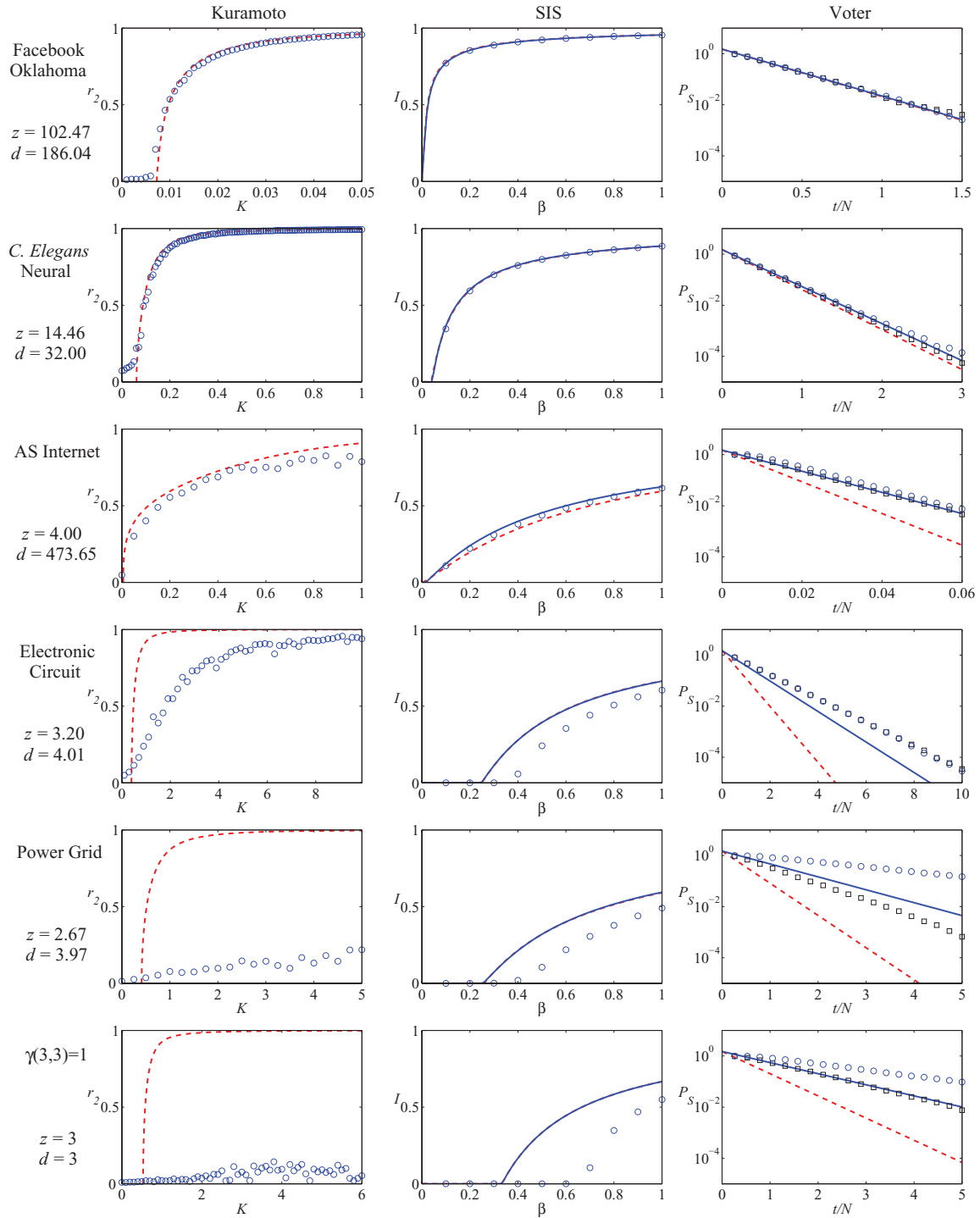
FIG. 4. (Color online) Results for dynamics of Figs. 1–3 for all six networks used in those figures. Curves and symbols are as in Figs. 1–3. For the voter model, black squares show the numerical results obtained by rewiring networks in a manner that conserves both degree distribution and degree-degree correlations [60].

In Fig. 5, we present results from numerical simulations of SIS dynamics and consider the final, steady-state fraction I of infected nodes. (For each network, we average over an ensemble of more than several hundred realizations; in each of these, 5% of the nodes are randomly chosen to be infected at $t = 0$.) Because the quality of the MF approximation is known to depend on the number of infected nodes [61], we further compare errors for different networks by choosing the

spreading rate β in each simulation so that the MF steady-state value I_{theory} equals 0.5. Letting

$$E_S = \frac{I_{\text{theory}} - I_{\text{numerical}}}{I_{\text{theory}}} \qquad (8)$$

be the relative error between the theoretical predictions of cMF theory (4,5) and the numerical results, we color each network's symbol in Fig. 5 by the value of E_S (see
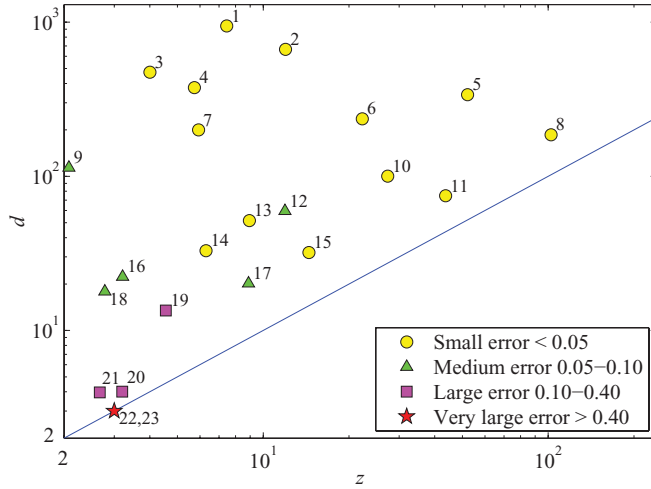
FIG. 5. (Color online) Accuracy of cMF theory for SIS dynamics as a function of values of mean degree z and mean nearest-neighbor degree d for many real-world networks. The numbers that label the networks are enumerated in column 1 of Table I. The colors indicate the magnitude of the relative error E_S, defined in Eq. (8), between cMF theory and numerical simulations for SIS dynamics with $I_{theory} = 0.5$.
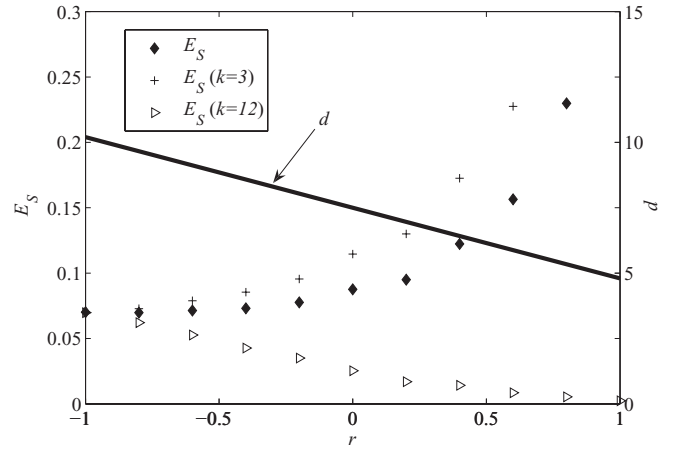

FIG. 6. Error (diamonds) of cMF theory in predicting steady-state infected fraction for SIS dynamics as a function of Pearson correlation coefficient r in the synthetic networks (described in the text) with $(k_1, k_2) = (3, 12)$. Also shown are the errors for the class of degree-3 nodes (crosses) and for the class of degree-12 nodes (triangles): these are calculated similar to Eq. (8), using $i_k(\infty)$ [from Eq. (4)] for I_{theory} and the average infected fraction of degree-k nodes in simulations for $I_{numerical}$. The solid line gives the mean nearest-neighbor degree d.

Table I for the error values). The results in Fig. 5 show clearly that situations with low d and low z correspond to inaccurate cMF predictions [see Figs. 1(b), 2(b), and 3(b)] and that the high-d situations (some of which also have small z) all have accurate cMF predictions, supporting our claim that the fidelity of MF theory can be evaluated using d.

In Sec. IV, we describe an alternative measure for predicting MF accuracy that uses intervertex distances [60]. One can also construct other, more complicated measures (e.g., by computing the size of the connected cluster of low-degree nodes); however, the mean first-neighbor degree d is appealing because it is simple to calculate and understand, and it has considerable explanatory power. We note that similarly accurate results have been found for real-world networks using MF theory for a discrete-time version of the SIS model [61,62].

B. Isolating the effects of dynamical correlations using synthetic networks

We have argued above that the observed accuracy of MF theory on some real-world networks is due to their high d values ameliorating the neglect of dynamical correlations [i.e., MF assumption (iii)]. However, real-world networks typically have high values for clustering coefficients and significant modular structures, so MF theory for such networks violates assumptions (i) and (ii) as well as assumption (iii). It might therefore be argued that the high-d effect seen in Fig. 5 could be due to an improvement in the validity of assumptions (i) and/or (ii), which could in principle have a larger impact than the high-d improvement of assumption (iii). We investigate this by now considering SIS dynamics on synthetic random networks (with $N = 10^4$ nodes) in which the transitivity and community structure are both negligible. Thus, MF assumptions (i) and (ii) are both valid for these networks, and the error in MF theory is due to violations of assumption (iii).

The first family of synthetic networks we use is described in Ref. [63]: Each node is either of low degree k_1 [with probability $p_{k_1} = k_2/(k_1 + k_2)$] or of high degree k_2 [with probability $p_{k_2} = k_1/(k_1 + k_2)$]. In order to create a network with a prescribed degree-degree correlation coefficient r, we connect the nodes of each type preferentially to nodes of either the same or of opposite type. We show in Fig. 6 how the aggregate error E_S (diamonds) depends on the correlation coefficient r for a specific case with $(k_1, k_2) = (3, 12)$. Note that the mean degree z of these networks is fixed ($z = 4.8$), but the mean first-neighbor degree d decreases as r increases. Figure 6 illustrates that the highest error for the MF theory occurs when d is lowest. This is the fully assortative ($r = 1$) case in which low-degree nodes link only to other low-degree nodes, creating a low-k connected cluster in which MF theory is inaccurate. At the other extreme, the disassortative ($r = -1$) case has every low-degree node linked only to high-degree nodes, with a consequent reduction in the error of MF theory (and a high value of d). The trend of the error in each degree class also supports our argument: high-degree neighbors correspond to lower MF error.

In Fig. 7, we show further details about the degree-3 nodes. To examine the importance of assumption (ii), we consider (for each value of r) an ensemble of $M = 25$ realizations of the SIS process on identical copies of the synthetic network. At a fixed (large) time, we record the state—either infected or susceptible—of each node i. Taking the average over the M realizations defines f_i, the fraction of realizations in which node i is infected at the chosen time. We now consider the distribution of f_i values over the set of nodes i that all share the same degree k, noting that assumption (ii) implies that these f_i values should be identical for every node i in a given degree class. Consequently, we plot in Fig. 7, for degree-3 nodes, the mean and standard deviation of the f_i values for the networks used in Fig. 6. Note that the mean value equals
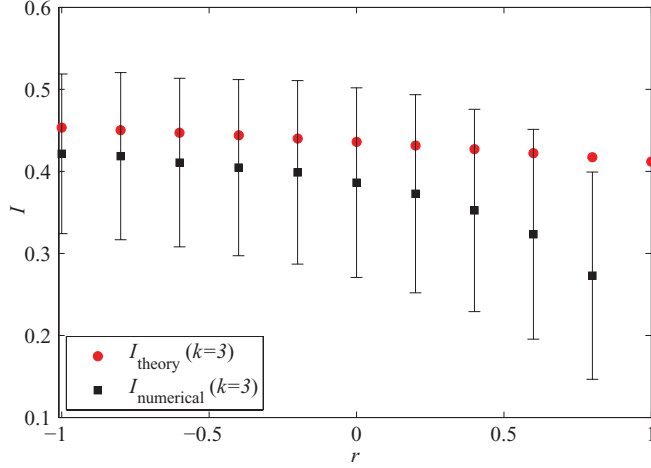
FIG. 7. (Color online) Mean (black squares) and standard deviation (error bars show one standard deviation above and below the mean) of the distribution of f_i values (described in text) on the same networks as in Fig. 6. The values of I_{theory} are given by cMF theory, using $i_3(\infty)$ from Eq. (4).

$I_{\text{numerical}}$ ($k = 3$), the average infected fraction of degree-3 nodes in simulations. This mean value deviates from the cMF theory prediction $i_3(\infty)$, giving the errors shown by the crosses in Fig. 6. We note that the standard deviation of the f_i values does not depend strongly on r (and hence also does not depend strongly on d). Moreover, if each of the f_i values for each r value were to be replaced by its mean value—so that the standard deviation for f_i was zero, in accord with assumption (ii)—then the errors shown in Fig. 6 would be unaltered. In this sense, the violation of assumption (ii) is not the main source of the high-d effect on the accuracy of MF theory.

To test for the existence of dynamical correlations—which are neglected under MF theory assumption (iii)—in the numerical results, we calculate the fraction ϕ_{SI} of SI edges (i.e., edges linking a susceptible node and an infected node) in the networks in steady state. We then compare this to the fraction of SI edges that would be obtained if no dynamical correlations were present. This is determined by treating as independent the probabilities for nodes at each end of an edge to be in states S and I:

$$\phi_{\text{SI}}^{\text{MF}} = 2 \sum_{k,k'} \rho_k \frac{k p_k}{z} P(k'|k)(1 - \rho_{k'}), \qquad (9)$$

where ρ_k is the fraction of infected degree-k nodes in the network [hence, $(1 - \rho_{k'})$ is the fraction of degree-k' nodes that are susceptible]. Note that we calculate ρ_k from the numerical simulations—we do not use the corresponding MF-theory values $i_k(\infty)$—so that we are directly testing the validity of assumption (iii) on the numerical data. The relative error of the MF assumption on the SI-edge fraction can then be calculated in a similar fashion to Eq. (8):

$$E_{\text{SI}} = \frac{\phi_{\text{SI}}^{\text{MF}} - \phi_{\text{SI}}}{\phi_{\text{SI}}^{\text{MF}}}. \qquad (10)$$

If assumption (iii) were true, there would be no dynamical correlation effects and the fraction of SI edges could be computed directly from the fraction of infected nodes using
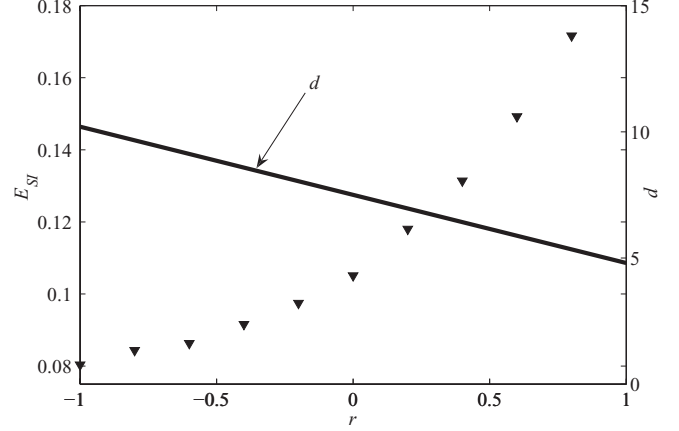

FIG. 8. Relative errors E_{SI} (symbols), given by Eq. (10), in the calculation of the fraction of SI edges using MF assumption (iii) on the same networks as in Fig. 6. The solid line gives the mean nearest-neighbor degree d.

Eq. (9), giving $E_{\text{SI}} = 0$. However, Fig. 8 shows that the error E_{SI} increases as the mean first-neighbor distance decreases, which is similar to the trend of error E_S in Fig. 6. This evidence supports our claim that it is MF assumption (iii)—the neglect of dynamical correlations—that plays a dominant role in determining the accuracy of MF theory on high-d networks.

The second family of synthetic networks is composed of networks with negligible degree-degree correlations and is generated using the configuration model [64]. In these networks, $d = \langle k^2 \rangle / z$, so the mean first-neighbor degree increases with the second moment of the degree distribution if the mean degree z is fixed. For example, one can construct networks with $z = 5$ with the degree probabilities

$$p_3 = \tfrac{15}{17}(1 - \alpha), \quad p_5 = \alpha, \quad p_{20} = \tfrac{2}{17}(1 - \alpha), \qquad (11)$$

and $p_k = 0$ for all other k. Such a network has a fraction α of nodes with degree 5, and the remaining nodes have degrees 3 and 20. The mean first-neighbor degree for such networks is $d = 11 - 6\alpha$, so it decreases linearly with α. By comparing numerical simulations for the SIS model with MF theory (not shown), we find that the error magnitude $|E_S|$ increases monotonically with α. It takes the value 0.07 at $\alpha = 0$ (with $d = 11$) and the value 0.16 at $\alpha = 1$ (with $d = 5$). Similar to the correlated synthetic networks of Fig. 6, the higher values of d thus correspond to lower values of the error.

The evidence from both families of synthetic networks suggests that the high-d effect that we have observed can be due only to its impact on dynamical correlations [i.e., assumption (iii)]. In real-world networks, d can presumably affect the validity of all three MF assumptions. Further work is required to understand which assumption(s) have the strongest impact on MF accuracy in such situations.

C. Mean-field accuracy for other dynamical processes

The argument that we have given above for the usefulness of d as a measure of MF accuracy is specific to SIS dynamics, where the quantity of interest is the (ensemble-averaged) infected fraction of nodes. We now consider error measures on the (z, d) plane for the Kuramoto and voter models. For the

Kuramoto model, we define an error measure in terms of the r_2 order parameter from Eq. (2) as

$$E_K = \frac{r_{2\,\text{theory}} - r_{2\,\text{numerical}}}{r_{2\,\text{theory}}}, \qquad (12)$$

which we evaluate at the value of K for which $r_{2\,\text{theory}} = 0.6$. Similarly, a measure of relative error for the voter model is given by

$$E_V = \frac{\log_{10}(P_{s\,\text{theory}}) - \log_{10}(P_{s\,\text{numerical}})}{\log_{10}(P_{s\,\text{theory}})}, \qquad (13)$$

where $P_{s\,\text{theory}}$ and $P_{s\,\text{numerical}}$ are the survival probabilities given by Eq. (6) and by numerical simulations, respectively. We evaluate these quantities at a time t that corresponds to a survival probability of 1% (in MF theory): $P_{s\,\text{theory}}(t) = 10^{-2}$. This definition reflects the vertical difference between the dashed curve and the symbols in Fig. 3 at a specific value of t.

We give the measured values for E_V and E_K for 21 real-world networks in Table I, and Figs. 9(a) and 9(b) show how these values depend on the mean degree z and mean first-neighbor degree d of the networks. The Kuramoto model exhibits a similar pattern to SIS dynamics [compare Fig. 9(a) with Fig. 5]: high-d networks have lower errors than low-d networks. However, the high-d effect does not seem to impact the voter model [see Fig. 9(b)] in the same way. We can understand this by contrasting the predicted quantities for SIS dynamics and the voter model. For SIS dynamics, the error is low when MF theory accurately predicts the fraction $I(t)$ of infected nodes. In the voter model, the quantity corresponding to $I(t)$ is the fraction of nodes in one of the two voter states, which we denote by $I_V(t)$. (By symmetry, $I_V(t)$ can be used to denote either of the two states without loss of generality.) When initial states are randomly assigned, MF theory predicts that $I_V(t)$ is conserved, which implies that $I_V(t) = 1/2$ for all t. Numerical simulations on real-world networks also give $I_V(t) = 1/2$, but only when one averages over an ensemble of realizations. In any single realization on a network of finite size, fluctuations eventually lead to the entire network becoming ordered—i.e., all nodes eventually share the same opinion. (In half of the realizations, this shared opinion is one voter state; in the other half, it is the other state.) It is this single-realization ordering process that is measured by the survival probability $P_s(t)$. In this respect, the voter model is different from both the SIS and Kuramoto models, in which the trajectory of a typical realization is qualitatively similar to an ensemble average of all trajectories.

In order to better capture quantities of higher order than $I_V(t)$, such as $P_s(t)$, it is necessary to approximate the dynamical correlations between nodes using, for example, a pair-approximation method. In Fig. 9(c), we show the magnitude of the PA error for the voter model; we still measure the error using Eq. (13), but we use the PA theory of Eq. (7) instead of the MF theory of Eq. (6). Observe the improvement in accuracy over the MF theory of Fig. 9(b), particularly for high-d networks. Similar to MF results for SIS dynamics (see Fig. 5) and for the Kuramoto model [see Fig. 9(a)], the largest errors are obtained in networks with both low z and low d.
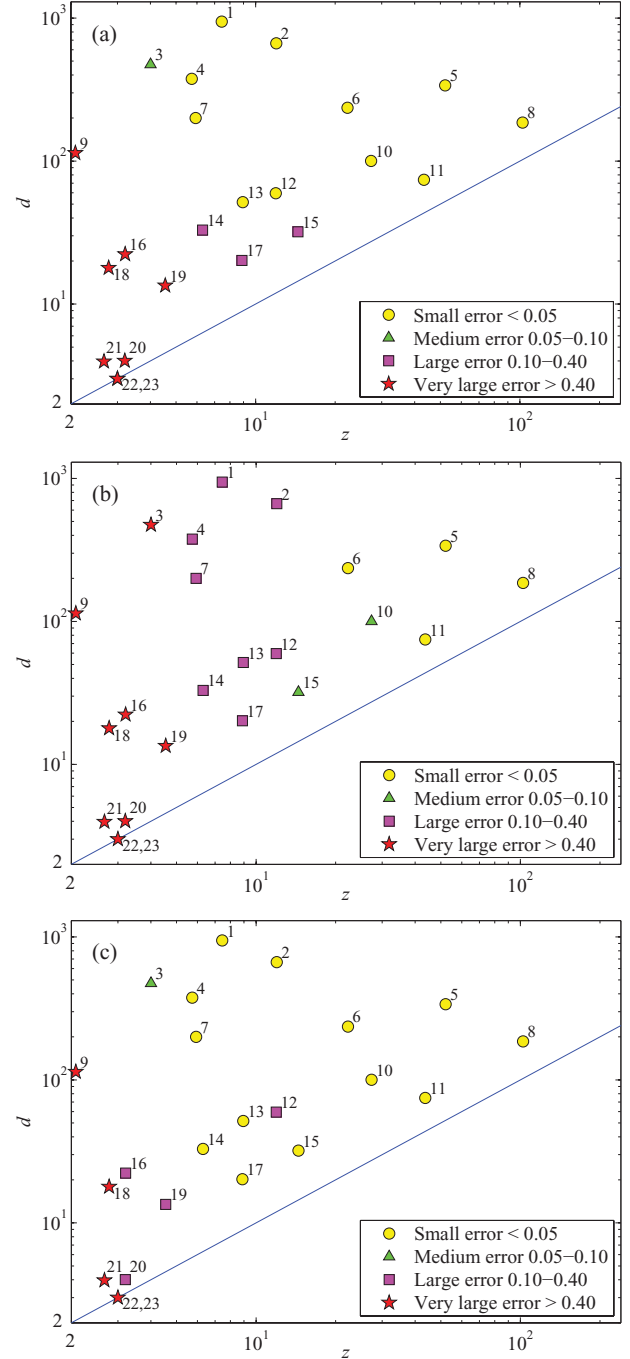

FIG. 9. (Color online) As in Fig. 5, except that color indicates the magnitude of the relative error, determined by comparing numerical results to (a) MF theory for the Kuramoto order parameter r_2 (with $r_{2\,\text{theory}} = 0.6$) from Eq. (12); (b) MF theory for the voter-model survival probability $P_s(t)$ at time t (with $P_{s\,\text{theory}} = 10^{-2}$) from Eq. (13); and (c) PA theory for the voter-model survival probability.

IV. COMPARISON WITH AN ALTERNATIVE MEASURE

Using numerous examples of real-world and synthetic networks, we have illustrated that the mean degree of first neighbors d is a good indicator of the accuracy of MF theories for a variety of dynamical processes on networks. One can also construct more complicated accuracy measures, which
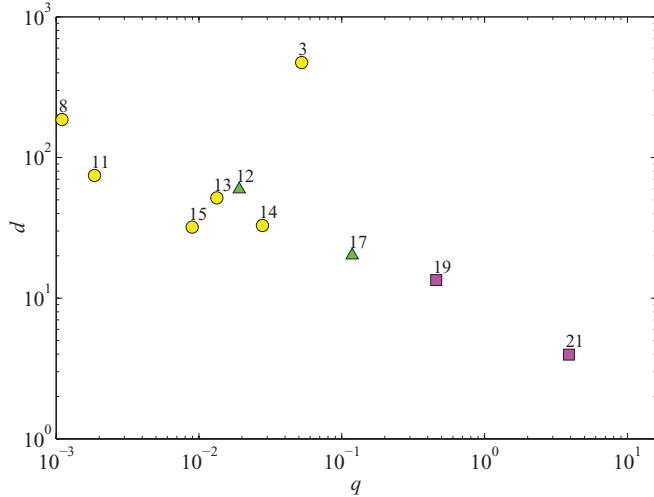
FIG. 10. (Color online) Location of real-world networks in the (q,d) parameter plane, where q is the error measure (14), which was shown in Ref. [60] to be correlated with the error of PA theory for bond percolation. We use the same symbols as in Fig. 5.

can in general depend on the dynamics under scrutiny. In Ref. [60], for example, we examined (among other dynamics) the accuracy of the bond percolation theory of [65] by comparing its predictions with numerical calculations of the sizes of the largest connected component for several real-world networks. We showed that a good measure of the error is given by the quantity

$$q = \frac{\ell - \ell_1}{z}, \qquad (14)$$

where ℓ is the mean intervertex distance in the original (clustered) network, ℓ_1 is the corresponding mean distance in a rewired version of the network (using a rewiring process that preserves degree-degree correlations but reduces clustering), and z is the mean degree. Noting that the bond percolation theory of Ref. [65] is of pair-approximation (PA) type, in contrast to the mean-field theories on which we focus in this paper, it is nevertheless of interest to examine the relation between q and the mean first-neighbor degree d that we have identified in this paper as an indicator of MF theory accuracy for several dynamical processes.

In the (q,d) parameter plane of Fig. 10, we show the positions of the real-world networks from Fig. 5 that were also examined in Ref. [60]. The expected relationship between q and d is revealed: Networks with high d have low q (and hence, according to Ref. [60], exhibit low error for bond percolation PA theory), and low d values correspond to high q and hence to large errors. Thus, despite its simplicity, d performs well when compared to other, more involved, measures. As noted above, there is scope for further work on developing more complicated

diagnostics (such as q) for predicting MF accuracy for various dynamics, but the use of the d value is appealing because it is simple to calculate and aids in understanding the underlying causes of MF inaccuracies.

V. CONCLUSIONS

In summary, we have shown that MF theory works best for networks in which low-degree nodes, if present, are connected to high-degree nodes (i.e., for networks that are either disassortative by degree or have high mean degree z). Remarkably, it is not necessary that the mean degree of a network be large for MF theory to work well—at least for ensemble-averaged quantities [see Figs. 5 and 9(a)]. In addition to the 21 real-world networks that we have studied, we have presented evidence from synthetic networks to support our hypothesis. We stress that our error measures focus on behavior far from critical points; the accuracy of MF theory for the calculation of phase-transition points (such as the value of K for the onset of synchronization in the Kuramoto model [66,67] or the SIS epidemic threshold [5,68–70]) is a topic for future work.

Based on our results for the voter model in Sec. III, we expect that similar conclusions should hold for the applicability of pair-approximation theories (such as those in Refs. [7,58]) for dynamics on real-world networks. Although PA theories account for the dynamical correlations that plague MF theories, they remain vulnerable to the effects of network clustering and modularity when d is low. For example, Fig. 3(b) gives an example in which PA theory works well only on the rewired (and hence unclustered) version of a (low-d) network. This suggests that PA theory (like MF theory) is most accurate for real-world networks with either high mean degree or high mean first-neighbor degree.

[1] A. Barrat, A. Vespignani, and M. Barthélemy, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).

[2] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[3] E. Estrada, Phys. Rev. E **84**, 047101 (2011).

[4] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, UK, 2010).

[5] R. Durrett, Proc. Natl. Acad. Sci. USA **107**, 4491 (2010).

[6] F. Vazquez and V. M. Eguíluz, New J. Phys. **10**, 063011 (2008).

[7] E. Pugliese and C. Castellano, Europhys. Lett. **88**, 58004 (2009).

[8] M. E. J. Newman, Phys. Rev. Lett. **103**, 058701 (2009).

[9] J. C. Miller, Phys. Rev. E **80**, 020901(R) (2009).

[10] J. P. Gleeson, Phys. Rev. E **80**, 036107 (2009).

[11] J. P. Gleeson, S. Melnik, and A. Hackett, Phys. Rev. E **81**, 066114 (2010).

[12] B. Karrer and M. E. J. Newman, Phys. Rev. E **82**, 066118 (2010).

[13] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[14] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).

[15] M. A. Porter, J.-P. Onnela, and P. J. Mucha, Not. Amer. Math. Soc. **56**, 1082 (2009), http://www.ams.org/notices/200909/rtx090901082p.pdf.

[16] S. N. Dorogovtsev et al., Rev. Mod. Phys. **80**, 1275 (2008).

[17] A. Baronchelli and V. Loreto, Phys. Rev. E **73**, 026103 (2006).

[18] A. Baronchelli and R. Pastor-Satorras, Phys. Rev. E **82**, 011111 (2010).

[19] A. L. Traud et al., SIAM Rev. **53**, 526 (2011).

[20] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[21] T. Ichinomiya, Phys. Rev. E **70**, 026116 (2004).

[22] Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence (Springer-Verlag, New York, 1984).

[23] A. Arenas et al., Phys. Rep. **469**, 93 (2008).

[24] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. E **71**, 036151 (2005).

[25] R. Milo et al., Science **303**, 1538 (2004).

[26] The CAIDA Autonomous System Relationships Dataset, 30-Jun-2008, http://www.caida.org/data/active/as-relationships; http://as-rank.caida.org/data/MAX2SAT/2008/as-rel.20080630.a0.01000.txt.

[27] R. Alberich et al., e-print arXiv:cond-mat/0202174.

[28] J. C. Johnson and L. Krempel, J. Soc. Str. **5**, 1 (2004), http://www.cmu.edu/joss/content/articles/volume5/JohnsonKrempel.

[29] A. L. Traud, P. J. Mucha, and M. A. Porter, Physica A (in press), e-print arXiv:1102.2166.

[30] K. Norlen et al., Proceedings of International Telecommunications Society 14th Biennial Conference (ITS2002), 2002.

[31] L. A. Adamic and N. Glance, LinkKDD ' 05: Proceedings of the 3rd International Workshop on Link Discovery, 2005, p. 36.

[32] V. Colizza et al., Nat. Phys. **3**, 276 (2007).

[33] A network obtained by considering the 500 US airports with the largest amount of traffic from publicly available data, http://sites.google.com/site/cxnets/US_largest500_airportnetwork.txt.

[34] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).

[35] A metabolic network of C. Elegans, http://deim.urv.cat/~aarenas/data/xarxes/celegans_metabolic.zip.

[36] V. Colizza et al., Physica A **352**, 1 (2005).

[37] V. Colizza et al., Nat. Phys. **2**, 110 (2006).

[38] Protein interaction network of the yeast Saccharomyces cerevisae extracted with different experimental techniques and collected at the Database of Interacting Proteins, http://dip.doe-mbi.ucla.edu/; http://sites.google.com/site/cxnets/DIP.dat.

[39] A neural network of C. Elegans, http://www-personal.umich.edu/~mejn/netdata/celegansneural.zip.

[40] R. Milo et al., Science **298**, 824 (2002).

[41] M. E. J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).

[42] Network of coauthorships between scientists posting preprints on the arXiv.org Condensed Matter E-Print Archive (which includes all preprints posted between 1-Jan-1995 and 31-Mar-2005), http://www-personal.umich.edu/~mejn/netdata/cond-mat-2005.zip.

[43] S. Mangan and U. Alon, Proc. Natl. Acad. Sci. USA **100**, 11980 (2003).

[44] X. Guardiola et al., e-print arXiv:cond-mat/0206240.

[45] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Phys. Rev. E **70**, 056122 (2004).

[46] Largest connected component of the network of users of the Pretty-Good-Privacy algorithm for secure information interchange, http://deim.urv.cat/~aarenas/data/xarxes/PGP.zip.

[47] An undirected, unweighted network representing the topology of the Western States Power Grid of the United States, http://www-personal.umich.edu/~mejn/netdata/power.zip.

[48] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[49] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[50] M. Barthélemy et al., J. Theor. Biol. **235**, 275 (2005).

[51] T. E. Harris, Ann. Probab. **2**, 969 (1974).

[52] R. M. Anderson and R. M. May, Infectious Diseases of Humans: Dynamics and Control (Oxford University Press, Oxford, 1992).

[53] F. Brauer and C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology (Springer-Verlag, New York, 2010).

[54] T. M. Liggett, Interacting Particle Systems (Springer, New York, 1985).

[55] V. Sood and S. Redner, Phys. Rev. Lett. **94**, 178701 (2005).

[56] R. Durrett, Random Graph Dynamics (Cambridge University Press, Cambridge, UK, 2007).

[57] S. A. Levin and R. Durrett, Philos. Trans. R. Soc. London B **351**, 1615 (1996).

[58] K. T. D. Eames and M. J. Keeling, Proc. Natl. Acad. Sci. USA **99**, 13330 (2002).

[59] T. House and M. J. Keeling, J. R. Soc. Interface **8**, 67 (2011).

[60] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Phys. Rev. E **83**, 036112 (2011).

[61] B. Guerra and J. Gómez-Gardeñes, Phys. Rev. E **82**, 035101(R) (2010).

[62] S. Gómez et al., Europhys. Lett. **89**, 38009 (2010).

[63] P. S. Dodds and J. L. Payne, Phys. Rev. E **79**, 066115 (2009).

[64] B. Bollobás, Random Graphs, 2nd ed. (Cambridge University Press, Cambridge, UK, 2001).

[65] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).

[66] H. Hong, M. Y. Choi, and B. J. Kim, Phys. Rev. E **65**, 026139 (2002).

[67] J. Gómez-Gardeñes, Y. Moreno, and A. Arenas, Phys. Rev. E **75**, 066106 (2007).

[68] C. Castellano and R. Pastor-Satorras, Phys. Rev. Lett. **105**, 218701 (2010).

[69] R. Parshani, S. Carmi, and S. Havlin, Phys. Rev. Lett. **104**, 258701 (2010).

[70] J. P. Gleeson, Phys. Rev. Lett. **107**, 068701 (2011).

[71] J.-P. Onnela et al., e-print arXiv:1006.5731.