The Unreasonable Effectiveness of Tree-Based Theory for Networks with Clustering

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

[1]Department of Mathematics & Statistics, University of Limerick, Ireland
[2]Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, OX1 3LB, UK
[3]CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK
[4]Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599-3250, USA
[5]Institute for Advanced Materials, Nanoscience & Technology, University of North Carolina, Chapel Hill, NC 27599-3216, USA

We demonstrate that a tree-based theory for various dynamical processes yields extremely accurate results for several networks with high levels of clustering. We find that such a theory works well as long as the mean intervertex distance ℓ is sufficiently small—i.e., as long as it is close to the value of ℓ in a random network with negligible clustering and the same degree-degree correlations. We confirm this hypothesis numerically using real-world networks from various domains and on several classes of synthetic clustered networks. We present analytical calculations that further support our claim that tree-based theories can be accurate for clustered networks provided that the networks are "sufficiently small" worlds.

PACS numbers: 89.75.Hc, 89.75.Fb, 64.60.aq, 87.23.Ge

I. INTRODUCTION

One of the most important areas of network science is the study of dynamical processes on networks [1–4]. On one hand, research on this topic has provided interesting theoretical challenges for physicists, mathematicians, and computer scientists. On the other hand, there is an increasing recognition of the need to improve the understanding of dynamical systems on networks to achieve advances in epidemic dynamics [5–7], traffic flow in both online and offline systems [8], oscillator synchronization [9], and more [3].

Analytical results for complex networks are rather rare, especially if one wants to study a dynamical system on a network topology that attempts to incorporate even minimal features of real-world networks. If one considers a dynamical system on a real-world network rather than on a grossly simplified caricature of it, then theoretical results become almost barren. Furthermore, most analyses assume that the network under study has a locally tree-like structure, so that they can only possess very few small cycles, whereas most real networks have significant clustering (and, in particular, possess numerous small cycles). This has motivated a wealth of recent research concerning analytical results on networks with clustering [7, 10–22].

Most existing theoretical results for (unweighted) networks are derived for an ensemble of networks using (i) only their degree distribution p_k, which gives the probability that a random node has degree k (i.e., has exactly k neighbors) or using (ii) their degree distribution and their degree-degree correlations, which are defined by the joint degree distribution $P(k, k')$ describing the probability that a random edge joins nodes of degree k and k'. In the rest of this paper, we will refer to case (i) as "p_k-theory" (the associated random graph ensemble is known as the "configuration model" [23]) and to case (ii) as "$P(k, k')$-theory". The clustering in sample networks is low in both situations; it typically decreases as N^{-1} as the number of nodes $N \to \infty$ [51].

We concentrate in this paper on undirected, unweighted real-world networks, which can be described completely using adjacency matrices. It is straightforward to calculate the empirical distributions p_k and $P(k, k')$, which can then be used as inputs to analytical theory for various well-studied processes. The results can subsequently be compared with large-scale numerical simulations using the original networks.

In the present paper, we demonstrate that analytical results derived using tree-based theory can be applied with high accuracy to certain networks despite their high levels of clustering. Examples of such networks include university social networks constructed using Facebook data [24] and the Autonomous Systems (AS) Internet graph [25]. Specifically, the analytical results for bond percolation, k-core sizes, and other processes accurately match simulations on a given (clustered) network provided that the mean intervertex distance in the network is sufficiently small—i.e., that it is close to its value in a randomly rewired version of the graph. Recalling that a clustered network with a low mean intervertex distance is said to have the *small-world property*, we find that tree-based analytical results are accurate for networks that are "sufficiently small" small worlds. In discussing this result, we focus considerable attention on quantifying what it means to be "sufficiently small".

The remainder of this paper is organized as follows. In Sec. II, we consider several dynamical processes on highly clustered networks and show that tree-based theory adequately describes them on certain networks but not on others. In order to explain our observations, we introduce in Sec. III a measure of prediction quality E and