Dense Percolation in Large-Scale Mean-Field Random Networks Is Provably “Explosive”

Alexander Veremyev^{1‡}, Vladimir Boginski^{1,2*}, Pavlo A. Krokhmal³, David E. Jeffcoat²

1 Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida, United States of America, **2** Research and Engineering Education Facility (REEF), University of Florida, Shalimar, Florida, United States of America, **3** Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, Iowa, United States of America

Abstract

Recent reports suggest that evolving large-scale networks exhibit “explosive percolation”: a large fraction of nodes suddenly becomes connected when sufficiently many links have formed in a network. This phase transition has been shown to be *continuous (second-order)* for most random network formation processes, including classical mean-field random networks and their modifications. We study a related yet different phenomenon referred to as *dense percolation*, which occurs when a network is already connected, but a large group of nodes must be *dense enough*, i.e., have at least a certain minimum required percentage of possible links, to form a “highly connected” cluster. Such clusters have been considered in various contexts, including the recently introduced *network modularity principle* in biological networks. We prove that, contrary to the traditionally defined percolation transition, dense percolation transition is *discontinuous (first-order)* under the classical mean-field network formation process (with no modifications); therefore, there is not only quantitative, but also qualitative difference between regular and dense percolation transitions. Moreover, the size of the largest dense (highly connected) cluster in a mean-field random network is explicitly characterized by rigorously proven tight asymptotic bounds, which turn out to naturally extend the previously derived formula for the size of the largest clique (a cluster with all possible links) in such a network. We also briefly discuss possible implications of the obtained mathematical results on studying first-order phase transitions in real-world linked systems.

Citation: Veremyev A, Boginski V, Krokhmal PA, Jeffcoat DE (2012) Dense Percolation in Large-Scale Mean-Field Random Networks Is Provably “Explosive”. PLoS ONE 7(12): e51883. doi:10.1371/journal.pone.0051883

Editor: Sergio Gómez, Universitat Rovira i Virgili, Spain

Received: June 26, 2012; **Accepted:** November 9, 2012; **Published:** December 18, 2012

Copyright: © 2012 Veremyev et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was supported by the following: Air Force Research Laboratory, Contract # FA8651-08-D-0108 <http://www.eglin.af.mil/units/afrlmunitionsdirector/index.aspx>; Air Force Office of Scientific Research, Grant # FA9550-12-1-0142 <http://www.afosr.af.mil/>; Defense Threat Reduction Agency, Grant # HDTRA1-09-1-0056 <http://www.dtra.mil/Research.aspx>; and National Science Foundation, Grant # EPS1101284 <http://www.nsf.gov>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: vb@ufl.edu

‡ Current address: NRC Research Associate at the Air Force Research Laboratory/Munitions Directorate, Eglin Air Force Base, Florida, United States of America

Introduction

In recent years, “explosive percolation” in large-scale random networks has received substantial attention. This phenomenon essentially means that as new links are randomly added step-by-step to a network with n nodes (assuming a very large n) and the number (or percentage) of links reaches a certain critical threshold (the phase transition point), the order of magnitude of the largest connected component changes abruptly from logarithmic in n to linear in n , which is often referred to as the emergence of a giant connected component.

Achlioptas et al. [1] recently reported an interesting computational study suggesting that although the percolation transition in the classical Erdős-Rényi (mean-field) random graph model (e.g., links being added to the network independently and with the same probability) is continuous, this transition can be made discontinuous (or, “explosive”) by applying modified network formation rules referred to as Achlioptas processes. Several related studies also addressed explosive percolation in various models [2–6]. However, other studies showed that explosive percolation is still continuous in most settings [7–9]. Recently, Riordan and Warnke [10,11] proved that explosive percolation is continuous for all

Achlioptas processes. As follows from these recent studies, the current state of knowledge on the classical percolation transition phenomena is that these phase transitions are continuous (*second-order*) under most network formation rules, but they can be discontinuous (*first-order*) if a network formation process is substantially different from the Erdős-Rényi model.

However, if one looks at the percolation transition phenomenon from the perspective of system *robustness*, percolation by itself does not necessarily ensure that a connected network, which has just undergone a percolation transition, is stable and resilient with respect to natural or man-made (possibly adversarial) impacts, which may disrupt multiple links and violate the overall integrity of the underlying system. In other words, the connectivity of the whole network or a large fraction of nodes (giant connected component) may not be sufficient for maintaining the “stability” of the linked system. For instance, intuitively, consider a group of particles linked by bonds: if the *relative number* of bonds is too small, this connected group of nodes would not form a stable structure (i.e., a crystal), since the destruction of a just a few bonds may disconnect the system. Note that a linear in n number of links ($n-1$ being the exact minimum) is sufficient to fully connect a system of n nodes. On the contrary, it is much harder to

disconnect a connected cluster containing close to the maximum possible number of links (i.e., $\gamma \times [n(n-1)/2]$ links, where $\gamma \in (0,1)$ is sufficiently close to 1). Such “dense” connected clusters will be a subject of special interest further in this paper.

“Dense” (or, “highly connected”) network clusters have recently received attention in diverse research areas, including those dealing with biological, social, communication, and financial networks, and, as it will be discussed below, these clusters may correspond to “functional units”, or “modules” in an underlying complex system, which is reflected in the so-called *network modularity* principle [12,13]. However, it turns out that finding large dense connected clusters in a network is a much more computationally challenging task than identifying “regular” connected components. Moreover, despite many previous results on the properties of connected components in random networks, dense clusters have not been studied nearly as much as “regular” connected components.

Therefore, in this work, we pursued a line of thought deviating from most of the recent percolation studies in random networks: instead of concentrating on various network formation rules (i.e., Erdős-Rényi vs. Achlioptas processes) and analyzing the traditional percolation transition phenomenon with the emerging giant connected component, we considered a different type of phase transition with the emphasis not just on connectivity, but on *dense connectivity*. We then rigorously analyzed how the *dense percolation transition* occurs under the classical Erdős-Rényi (mean-field) network evolution process, as well as how the largest dense cluster size behaves when the whole network is sparse.

Results

Dense Clusters in Networks

To facilitate the description of the main results, we first introduce basic definitions and briefly discuss the relationship of dense connected clusters and “regular” connected components in networks. A *dense connected component* (cluster) in a network is defined in terms of its *edge density* parameter ρ , which is equal to the ratio of the actual number of links (edges) to the maximum possible number of links within the cluster (in other words, the relative percentage of links within the cluster – see Fig. 1 for illustration).

In the basic graph-theoretic model, the minimum required edge density for a cluster to be dense enough (“stable”) does not depend on the size of the considered cluster and is set to a fixed parameter $\gamma \in (0,1]$, which may be chosen depending on application-specific considerations (that is, a cluster is considered dense enough, if $\rho \geq \gamma$). Note that although the minimum required edge density γ is fixed, the number of links required to form a dense cluster on q nodes does depend on the size of the cluster q and is equal to $\gamma[q(q-1)/2]$. The concept of dense clusters was also recently addressed in the context of *network modularity* principle described by Hartwell et al. [12] and further analyzed by Spirin and Mirny [13], who considered “*highly connected subgraphs (clusters)*” (or, “modules”) defined using the aforementioned edge density parameter, which represented meaningful functional units in molecular networks. The significance of dense clusters in other research areas will be addressed in the next section.

In graph theory, dense (“highly connected”) clusters described above are referred to as γ -*quasi-cliques* [14,15], with the limiting case of $\gamma=1$ corresponding to the well-known *clique* structure, where there is a link between *every* two nodes. Clearly, cliques are very cohesive and resilient to node and link failures; however, large-scale real-world systems rarely form cliques, as it is often unrealistic to connect every two nodes by a link. Therefore, quasi-cliques provide a reasonable tradeoff between regular connected components, which do not necessarily reflect network modularity and may be “not robust enough” (as indicated above), and cliques, which are “too robust”. The assumption that the required threshold value of γ for a cluster with q nodes to form a quasi-clique (“stable”/“highly connected” cluster) is constant and does not depend on q may not always be fully realistic; however, it does make theoretical and practical sense, since if one considers the large-scale (asymptotic) case with $q \rightarrow \infty$, the function $\gamma(q)$ should not vanish as q increases, since otherwise this would represent an arbitrarily low edge density of a cluster, which would effectively “downgrade” a dense connected cluster to a regular connected component.

In related work, the concept of a k -core (a connected subgraph where each node has a degree of at least k , that is, at least k neighbors) has been analyzed in mean-field random networks [16–18], and D’Souza [19] discussed the notion of “dense k -cores” as

Figure 1. Small-scale illustration of the concept of dense connected clusters in comparison with traditionally defined connected components. (A) A connected network with 20 nodes and 22 links. Assuming that the minimum required percentage of links for a group of nodes to form a dense connected (“highly connected”) cluster is $\gamma = 70\%$ (this value is chosen arbitrarily for illustrative purposes only), the largest-size dense cluster is 4, and all other dense clusters have only 2 nodes each. (B) The network from (A) after 10 more links have formed (for instance, this may be the result of increasing the value of p if one assumes $G(n,p)$ model). Dense clusters with the largest number of nodes are highlighted, with their size still significantly smaller than the size of the whole network. It turns out that if n is very large, further increase of p in $G(n,p)$ will only produce dense clusters scaling as $\log(n)$ (no clusters scaling linearly with n) until p reaches the critical point $p_* = \gamma$, after which the whole network abruptly becomes a dense cluster.

doi:10.1371/journal.pone.0051883.g001

clusters that may possess extra robustness properties in addition to connectivity. However, if k is fixed and $n \rightarrow \infty$ there exist k -cores with edge density arbitrarily close to zero. Therefore, a more meaningful type of a “robust cluster”, which preserves the required edge density in addition to the minimum required degree of each node, is a slightly modified version of a k -core, where k is defined in terms of the *percentage* of all possible neighbors within a cluster, i.e., $k = \gamma(q-1)$ [20–22], that is, the number of neighbors for each node within the cluster *depends on the size of this cluster*. It is easy to observe that any $\gamma(q-1)$ -core is also a γ -quasi-clique, but the reverse is not always true. Therefore, $\gamma(q-1)$ -cores represent a slightly more restrictive type of “highly connected” clusters than γ -quasi-cliques. In addition, $\gamma(q-1)$ -cores possess certain guaranteed robustness properties if $\gamma > 1/2$: as it straightforwardly follows from [23], a $\gamma(q-1)$ -core of size q with $\gamma > 1/2$ has a diameter of at most 2 (that is, any two nodes are connected through at most one intermediary), and it would stay connected after the removal of up to $(2\gamma-1)(q-1)$ links.

Under the aforementioned assumption of the fixed required value of γ , we posed a natural question: how does the size of the largest dense connected cluster (γ -quasi-clique or $\gamma(q-1)$ -core) grow as more and more links are added to an evolving large-scale network? In particular, can one rigorously prove the emergence of a “giant dense connected component” and investigate the order of this phase transition (i.e., whether it is continuous or discontinuous)? It turns out that in the context of the classical $G(n, p)$ (Erdős-Rényi) model, these questions can be answered unambiguously.

First-Order Dense Percolation Transition in Erdős-Rényi Networks

Formally, an Erdős-Rényi random graph $G(n, p)$ contains n nodes, and each pair of nodes is connected by a link independently with probability p . The process of evolution of such a network can be represented simply by a gradual increase of the parameter p , with larger values of p representing more links in the network.

Table 1. Maximum γ -quasi-clique sizes in $G(n, p)$ for $n = 100$, and $\gamma = 0.85, 0.90$.

p	$\gamma = 0.85$			$\gamma = 0.90$		
	$[m_n^l, m_n^u]$	$\mathbb{E}[M_n^\gamma]$	$[\underline{M}_n^\gamma, \bar{M}_n^\gamma]$	$[m_n^l, m_n^u]$	$\mathbb{E}[M_n^\gamma]$	$[\underline{M}_n^\gamma, \bar{M}_n^\gamma]$
0.05	[3.07, 4.32]	3.10	[3, 4]	[3.07, 3.88]	3.10	[3, 4]
0.06	[3.27, 4.66]	3.26	[3, 5]	[3.27, 4.16]	3.26	[3, 5]
0.07	[3.46, 4.98]	3.38	[3, 5]	[3.46, 4.44]	3.38	[3, 5]
0.08	[3.65, 5.30]	3.60	[3, 5]	[3.65, 4.71]	3.60	[3, 5]
0.09	[3.82, 5.62]	3.97	[3, 5]	[3.82, 4.97]	4.97	[3, 5]
0.10	[4.00, 5.94]	4.27	[3, 5]	[4.00, 5.50]	4.24	[3, 5]
0.11	[4.17, 6.26]	4.66	[4, 5]	[4.17, 5.77]	4.65	[4, 5]
0.12	[4.34, 6.58]	4.87	[4, 6]	[4.34, 6.04]	4.82	[4, 5]
0.13	[4.51, 6.91]	5.06	[5, 6]	[4.51, 6.04]	5.00	[5, 5]
0.14	[4.68, 7.25]	5.19	[5, 6]	[4.68, 6.31]	5.01	[5, 6]
0.15	[4.85, 7.59]	5.67	[5, 7]	[4.85, 6.59]	5.06	[5, 6]

Average ($\mathbb{E}[M_n^\gamma]$), minimum ($\underline{M}_n^\gamma$), and maximum ($\bar{M}_n^\gamma$) observed values of the largest γ -quasi-clique size, computed for 100 instances of uniform random graphs $G(n, p)$ for $n = 100$ and $\gamma = 0.85, 0.90$. Theoretical lower and upper bounds $[m_n^l, m_n^u]$ are calculated using formula (29).
doi:10.1371/journal.pone.0051883.t001

Although this is a somewhat idealized classical model, the issues of its appropriateness in certain contexts will be addressed in the next section.

One may intuitively assume that the size of the largest dense connected component in $G(n, p)$ would first grow logarithmically (while p is much smaller than γ) and then linearly right after some critical point $p_c < \gamma$, and it would continue to grow until it eventually reaches n , similarly to the asymptotic behavior of the giant connected component in an Erdős-Rényi network; however, our rigorous mathematical arguments show that this is not the case. As our results show, for $n \rightarrow \infty$, the largest dense connected cluster scales as $\log(n)$ as long as $p < \gamma$ (even if p is very close to γ), whereas when p reaches γ , the whole network abruptly becomes a dense connected cluster. Essentially, for very large values of n , the jump from the logarithmic order of magnitude of the largest dense cluster to the largest dense cluster of size n occurs at one point $p_* = \gamma$, with no “linear growth” phase in between! Therefore, for any fixed γ , there is no emerging giant dense connected component as p gradually increases, until a discontinuous jump at the point p_* produces a dense connected component representing the entire network. In other words, the size of the largest dense connected component in a large-scale Erdős-Rényi random network exhibits a *first-order phase transition*; moreover, the existence of this phase transition has been proven by fully rigorous analytical arguments.

The complete detailed proofs are presented in the *Materials and Methods* section, and here we briefly summarize the obtained formal results. The following proven facts characterize the size of the largest dense connected component (γ -quasi-clique) in $G(n, p)$ for sufficiently large n .

If M_n^γ is the size of the maximum γ -quasi-clique in $G(n, p)$ for some fixed $\gamma \in (0, 1)$, then for any $p < \gamma$ it holds *almost surely* (a.s.) that

$$M_n^\gamma \in \left[\frac{2 \ln n}{\ln \left(\frac{1}{p} \right)}, \frac{2 \ln n}{\ln \left[\left(\frac{\gamma}{p} \right)^\gamma \left(\frac{1-\gamma}{1-p} \right)^{1-\gamma} \right]} \right] \quad \text{a.s.} \quad (1)$$

In this context, a property $\mathcal{Q}_n$ is said to hold *almost surely* (a.s.), if with probability 1 there exists n_0 such that $\mathcal{Q}_n$ holds for all $n \geq n_0$.

Formula (1) is also valid for the size of the largest $\gamma(q-1)$ -core (in the context of the above description) in $G(n, p)$. Moreover, *with high probability* (w.h.p.)

$$\lim_{p \nearrow \gamma} \lim_{n \rightarrow \infty} \frac{M_n^\gamma}{n} = 0, \quad (2a)$$

but

$$\lim_{p \searrow \gamma} \lim_{n \rightarrow \infty} \frac{M_n^\gamma}{n} = 1, \quad (2b)$$

where a property $\mathcal{Q}_n$ is said to be observed with high probability (w.h.p.) if the probability of $\mathcal{Q}_n$ being observed converges to 1 as $n \rightarrow \infty$.

Expression (1) entails that the size of the largest γ -quasi-clique does scale logarithmically with n for any $p < \gamma$. Note that it provides not only the order of magnitude, but also asymptotically precise upper and lower bounds on the size of the largest γ -quasi-clique. An interesting observation here is the fact that the upper bound converges to the lower bound as $\gamma \rightarrow 1$, whereby (1) becomes the classical asymptotic estimate for the maximum clique size in

Figure 2. Illustration of the behavior of the largest dense cluster size in moderate-size graphs $G(n, p)$. (A) Relative size of maximum γ -quasi-cliques (M_n^γ/n) in random graphs $G(n, p)$ for $\gamma = 0.5$ (this particular value is chosen simply for illustrative purposes: the obtained results hold for any fixed $\gamma \in (0, 1)$) and $n = 500, 1,000, 5,000, 10,000, 20,000$. Due to the fact that finding the maximum quasi-clique in a graph is a computationally challenging NP-hard problem (as opposed to finding the largest “regular” connected component, which can be done in polynomial time), the numerical simulations were carried out using GRASP heuristic algorithm [14] and plotting the largest relative size found after multiple runs of the algorithm for each p . The growth increment of p was chosen at $\Delta p = 0.001$ in the region where p approaches γ from below (more details on computational experiments are given in Materials and Methods). (B) Theoretical behavior of the relative size of the maximum γ -quasi-clique in $G(n, p)$ as $n \rightarrow \infty$, which is simply a step function, as indicated by formulas (2a)–(2b). doi:10.1371/journal.pone.0051883.g002

$G(n, p)$ described by Bollobás and Erdős [24] and Grimmett and McDiarmid [25].

Formulas (2a) and (2b) formally express the existence of the first-order (discontinuous) jump of the size of the largest dense connected component (M_n^γ) compared to the size of the whole network (n): if $p < \gamma$ and p approaches γ from below (that is, the edge density of the whole network is just below the required edge density γ), the size of the largest cluster that does have the required density γ is still negligible compared to the size of the whole network. For $p > \gamma$ and approaching γ from above, the whole network forms the dense cluster, and as mentioned above, there is no “gradual” (continuous) change in terms of the size of the largest dense cluster.

As a final remark, the fact that the location of the dense percolation transition is in the point $p_* = \gamma$ is not surprising and can be intuitively understood, since one would expect that the size of the largest dense connected cluster in $G(n, p)$ would be equal to n if $p > \gamma$ and less than n if $p < \gamma$. However, the fact that this phase transition is *first-order in the asymptotic case* is not intuitively

predictable. This fact has been established via advanced analytical arguments that will be presented below.

Discussion

In this section, we briefly discuss potential implications of the obtained results on studying network clusters and phase transitions in related research areas. Dense connected clusters (γ -quasi-cliques, $\gamma(q-1)$ -cores and similar models) have been recently utilized to study large-scale networked systems in a variety of disciplines, including biological networks [12,13,21,26–29], social networks [30–33], telecommunication networks [14,15], and financial networks [34–37]. On the other hand, the mechanisms of phase transitions in large-scale physical systems are not completely understood, with a number of open questions remaining. Although mean-field random networks have certain limitations in modeling real-world systems, $G(n, p)$ -based network models have been employed in the literature to study various physical processes, including NMR sequential assignment [38],

Figure 3. Illustration of the behavior of the largest dense cluster size in moderate-size graphs $G(n, p)$. Relative size of the maximum γ -quasi-cliques in the uniform random graphs for $\gamma=0.7$, $n=500, 1,000, 5,000, 10,000, 20,000$, and $0 < p < 0.8$. All notations are the same as in the caption of Fig. 2.

doi:10.1371/journal.pone.0051883.g003

Potts glass formation [39], and relation of $G(n, p)$ to quantum theory [40].

In general, according to our results, natural systems whose evolution can (to some extent) be described by a mean-field random network model would exhibit a first-order dense percolation transition, if the size n of the system is *very large* (e.g., a mole of liquid contains $n = 6.02 \times 10^{23}$ particles, which for most purposes can be considered $n \rightarrow \infty$). For instance, if one assumes that the process of bond formation in a large-scale system of particles has a purely probabilistic nature (i.e., indistinguishable particles moving randomly with respect to each other, allowing one to make a simplifying assumption that the formation of a link/bond between any two particles *occurs with the same probability*), then a “highly connected” cluster spanning the whole system (with the required edge density γ , which may be set depending on characteristics of a particular system) would emerge abruptly when the probability of link formation p reaches γ . However, it should be noted that many real-life networks are not dense; therefore, examples of dense percolation transition phenomena with γ close to 1 cannot be easily identified in nature.

In addition, formula (1) has important implications in computational studies of the maximum γ -quasi-clique problem. As it turns out, identifying the largest γ -quasi-clique in a given network is an extremely computationally challenging task. The recent work [41] shows that this problem is NP-hard, which means that finding its exact solution for a network of size n would require an exponential in n number of operations. The currently available

exact methods allow one to explicitly compute the maximum γ -quasi-clique size only in small sparse graphs ($n \sim 10^2$) [41]. Therefore, the performance of any inexact (approximate/heuristic) methods cannot be evaluated for *large-scale* networks. Thus, the theoretical bounds for the maximum γ -quasi-clique size in large-scale mean-field networks are of particular interest in the context of evaluating the performance of new computational algorithms that may be developed for this problem in future.

Overall, the results of this study provide a starting point for rigorous mathematical justifications for the existence of first-order phase transitions in large-scale linked systems. Despite the limitations of classical mean-field random networks, the process of network evolution under this model may capture certain aspects of the evolution of natural systems undergoing phase transitions. Moreover, due to the fact that dense percolation transition is already discontinuous without any changes to the mean-field model (as opposed to “regular” percolation that can be made discontinuous only after substantial modifications to the mean-field model), one can hypothesize that dense percolation may in fact be discontinuous for other network formation rules. Therefore, dense percolation transition can potentially be further analyzed in the context of other random network formation processes reflecting the behavior of physical, biological, and social systems.

Materials and Methods

In this section we present graph-theoretic definitions and rigorous proofs of the aforementioned results. Since the nature of this study is theoretical rather than experimental, the main emphasis is put on mathematical proofs, rather than on data collection description.

Our formal arguments establish the existence of the first-order dense percolation transition in random graphs in the *asymptotic* ($n \rightarrow \infty$) sense, i.e., for very large graphs. Note, however, that the mathematical techniques that support this result cannot be employed to draw conclusions on the situation with smaller networks. In view of that, we complement our theoretical findings by computational studies that illustrate the behavior of maximum γ -quasi-cliques in small- to moderately-sized graphs. Note also that the computational cost of conducting such experiments quickly becomes prohibitive as the size of the network increases: this is a direct consequence of the aforementioned fact that the maximum γ -quasi-clique problem is NP-hard [41].

Due to the fact that the paper mainly considers the mean-field (Erdős-Rényi) random graph model, the networks utilized in numerical experiments were randomly generated using a straightforward procedure: for each value of p , the link between each pair of nodes was generated randomly and independently with probability p .

Graph-Theoretic Notations, Definitions, Related Work

A simple undirected graph $G=(V,E)$ is defined in terms of its set of vertices (nodes) V ($|V|=n$) and its set of edges (links) E . A *complete* subgraph, or a *clique* in a graph $G=(V,E)$ is a subset $Q \subseteq V$ of vertices that are pairwise adjacent, i.e., for any $i,j \in Q$ there exists an edge $(i,j) \in E$. In the present work we are concerned with the properties and behavior of a certain type of *clique relaxation*, namely the γ -quasi-clique [14].

Definition 1 (γ -quasi-clique, or γ -dense subgraph) *Let $G=(V,E)$ be a simple undirected graph, and $Q \subseteq V$ be a subset of its vertices. The (induced) subgraph $G[Q]$ is called a γ -quasi-clique for a given fixed $\gamma \in (0,1]$, if the ratio of the number of edges in $G[Q]$ to the maximum possible number of edges among vertices in Q is at least γ :*

$$|E(Q)| \geq \gamma \binom{|Q|}{2}.$$

Note that the case of $\gamma=0$ would be trivial, as any graph is a 0-quasi-clique. A complete graph is a 1-quasi-clique, hence γ -quasi-clique represents a *density*-based relaxation of the clique, as compared to *degree*- and *diameter*-based clique relaxations such as k -plex and k -club [42–50]. A γ -quasi-clique with the largest number of vertices is called the *maximum γ -quasi-clique*.

In this work, we investigate the asymptotic behavior of γ -quasi-cliques in large-scale random graphs. In particular, we employ the well-known $G(n,p)$ model of random graphs, originated by Erdős [51], which denotes a graph on n vertices, such that an edge between any two vertices exists with a probability $p \in (0,1]$, independently from other edges. Since the $G(n,p)$ model yields graph instances with a rather “uniform” structure, as opposed to, for instance, *power-law* graphs, it is often called a *uniform random graph* model [52].

Random graphs and related structures, such maximum cliques in random graphs, have been studied intensively in last decades [24,53,54]. One of the earliest works on the asymptotic behavior of maximum clique in uniform random graphs is due to [55], who showed that the maximum clique size has a strong peak around

$2 \ln n / \ln(1/p)$. Grimmett and McDiarmid [25] proved that as $n \rightarrow \infty$ the maximum clique size in a uniform random graph $G(n,p)$ is equal to $2 \ln n / \ln(1/p) + O(\ln \ln n)$ with probability one.

Main Theorems and Remarks

First, we prove a generalization of the result of Grimmett and McDiarmid [25] for the case of maximum γ -quasi-clique size. Furthermore, we demonstrate that the size of maximum γ -quasi-clique in $G(n,p)$ undergoes a *phase transition* when the value of p is varied in the vicinity of the (fixed) value $\gamma \in (0,1)$, manifested in a sudden and drastic change of size of the maximum γ -quasi-clique in $G(n,p)$ relative to the size of the graph itself. Specifically, in the next subsection we present the proofs of the following two theorems.

Theorem 1 *If $0 < p < \gamma \leq 1$, then the size M_n^γ of the maximum γ -quasi-clique in a uniform random graph $G(n,p)$ satisfies*

$$\frac{2}{\ln 1/p} \leq \liminf_{n \rightarrow \infty} \frac{M_n^\gamma}{\ln n} \leq \limsup_{n \rightarrow \infty} \frac{M_n^\gamma}{\ln n} \leq \frac{2}{\ln \left[\left(\frac{\gamma}{p} \right)^\gamma \left(\frac{1-\gamma}{1-p} \right)^{1-\gamma} \right]} \quad a.s. \quad (3)$$

Theorem 2 *If M_n^γ is the size of the maximum γ -quasi-clique in a uniform random graph $G(n,p)$ for some fixed $\gamma \in (0,1)$, then with high probability (w.h.p.)*

$$\lim_{p \nearrow \gamma} \lim_{n \rightarrow \infty} \frac{M_n^\gamma}{n} = 0, \quad (4a)$$

but

$$\lim_{p \searrow \gamma} \lim_{n \rightarrow \infty} \frac{M_n^\gamma}{n} = 1. \quad (4b)$$

Second, in addition to the computational experiments on relatively large graphs, where large γ -quasi-cliques were found using heuristic algorithms (due to NP-hardness of the maximum γ -quasi-clique problem for any fixed $\gamma \in (0,1]$, as stated above), we also conducted experiments on smaller graphs using a linear mixed integer programming formulation for the maximum quasi-clique problem, which allowed us to identify *exact* maximum quasi-cliques in graphs with up to 100 vertices. Interestingly, it turned out that the obtained asymptotic bounds were rather accurate for the considered small-scale uniform random graph instances. The details of these computational experiments are presented further in this section.

Remark 1 Note that while in the context of this work we are interested in the largest *dense connected component*, the above definition of γ -quasi-clique does not require connectivity. This allows for significant simplifications in the arguments that are presented below; however, the obtained results are still valid for the largest *connected γ -quasi-clique* due to the following observations:

1. It can be easily shown that if $p > \gamma$, then the whole graph $G(n,p)$ is w.h.p. a γ -quasi-clique. Since under the model assumptions $\gamma \in (0,1]$ is a parameter that does not depend on the

size of the graph n , then if $p > \gamma$, the whole graph $G(n, p)$ is automatically connected w.h.p., as a direct application of classical results by Erdős and Rényi [53]; therefore, the connectivity requirement for the largest γ -quasi-clique is w.h.p. satisfied in this case.

2. If the largest γ -quasi-clique does not coincide with the whole graph $G(n, p)$ (this would correspond to the case $p < \gamma$), asymptotically precise upper and lower bounds (1) that will be proven for the size of the largest γ -quasi-clique (in the context of Definition 1) are automatically valid for the largest *connected* γ -quasi-clique. This is due to the simple observation that the size of the largest connected γ -quasi-clique does not exceed the size of the largest (not necessarily connected) γ -quasi-clique, and it is at least as large as the size of the maximum clique.

Remark 2 The phase transition, a phenomenon of a drastic change in some property of a random structure over a small change in the structure's parameters, is well known in the literature. With respect to random graphs, the limiting probability of a graph's property changing from 0 to 1 or vice versa is well known for *monotone* and *first order* graph properties [56]. A property $\mathcal{Q}$ is *monotone increasing* (respectively, *decreasing*) if from $A \subseteq B$ (resp., $B \subseteq A$) and $A \in \mathcal{Q}$ it follows that $B \in \mathcal{Q}$. The first order graph properties are ones that can be finitely described in a first order language, i.e., language consisting of variables that represent graph vertices, equality ($=$) and adjacency ($\sim$) relations, Boolean symbols $\vee, \wedge, \neg$, and the universal and existential quantifications $\forall, \exists$. Note that first order properties are not necessarily monotone and vice versa; for instance, the increasing property “graph is connected” cannot be expressed in first order language [57]. Then, limiting relations similar to (23) that concern random graphs with first order properties $\mathcal{Q}$ are known as *zero-one laws* [56,57]:

$$\lim_{n \rightarrow \infty} \mathbb{P}\{G(n, p) \text{ has } \mathcal{Q}\} = 0 \text{ or } 1,$$

where the probability is monotone if $\mathcal{Q}$ is monotone.

In this context, it is worth noting that the property that “graph is a γ -quasi-clique” in neither monotone, nor first order property, hence the phase transition in the relative size of γ -quasi-clique in uniform random graphs (23) *may not be obtained directly from the general zero-one laws relations*.

Remark 3 Another related definition of the Erdős-Rényi random graph model that exists in the literature is $G(n, M)$ where graphs on n nodes with M links are uniformly equiprobable with

probability $\left(\frac{\binom{n}{2}}{M}\right)^{-1}$. In this work, we adhere to the $G(n, p)$

definition given above. All the presented results consider $p \in (0, 1)$ explicitly depending on γ rather than on n (that is, for sufficiently large n , the condition $p \gg \ln n/n$ is always true), so the *whole graph* is with high probability (w.h.p.) connected according to the classical theory [53]. Although the largest connected component already coincides with the whole graph, the largest *dense connected component* may still be small compared to the size of the whole graph, which turns out to be the case for any $p < \gamma$.

Remark 4 The obtained asymptotic bounds and first-order phase transition results are also valid for the size of the maximum $\gamma(q-1)$ -core (k -core with $k = \gamma(q-1)$ mentioned above), or, more generally, for the size of the largest subgraph that is required to have a certain minimum degree $\lambda > 0$ as a *percentage* of its size, in addition to the minimum edge density γ . These subgraphs (clusters) are referred to as (λ, γ) -quasi-cliques [20]. Formally, a

subgraph of $G = (V, E)$ induced by the set of vertices $V' \subset V$ is a (λ, γ) -quasi-clique ($1 \geq \gamma \geq \lambda \geq 0$) if and only if the following two conditions hold:

$$\forall v \in V' : \deg_{V'}(v) \geq \lambda(|V'| - 1), \quad (5)$$

$$|E'| \geq \gamma \binom{|V'|}{2}, \quad (6)$$

where $E' = E \cap (V' \times V')$ and $\deg_{V'}(v)$ is the number of neighbors of vertex v in set V' .

Let $M_n^{(\lambda, \gamma)}$ be the size of the maximum (λ, γ) -quasi-clique in random graph $G(n, p)$. Observe that $M_n^1 \leq M_n^{(\lambda, \gamma)} \leq M_n^\gamma$, i.e., the maximum size of (λ, γ) -quasi-clique is always not greater than the maximum size of γ -quasi-clique and bounded by the maximum clique size. It follows from the fact that any clique is a (λ, γ) -quasi-clique, and any (λ, γ) -quasi-clique is a γ -quasi-clique.

Therefore, the bounds in (21) are also valid for $M_n^{(\lambda, \gamma)}$. Moreover, it can be easily shown that when $p > \gamma$

$$\mathbb{P}\{M_n^{(\lambda, \gamma)} = n\} \rightarrow 1, \quad n \rightarrow \infty,$$

thus, the *first-order* phase transition in the point $p^* = \gamma$ is also valid for the asymptotic behavior of the order of magnitude of the maximum (λ, γ) -quasi-clique (including $\gamma(q-1)$ -core as a special case).

Rigorous Proofs of Main Results

Define N_k^γ as the (random) number of γ -quasi-cliques of size k in $G(n, p)$. Noting that there are $\binom{n}{k}$ different subgraphs of size k in this graph, let I_j^γ be the indicator variable such that $I_j^\gamma = 1$ if j -th subgraph of size k is a γ -quasi-clique, $j = 1, \dots, \binom{n}{k}$, and $I_j^\gamma = 0$ otherwise, which allows us to express N_k^γ as

$$N_k^\gamma = \sum_{j=1}^{\binom{n}{k}} I_j^\gamma. \quad (7)$$

Obviously, the unconditional probabilities $\mathbb{P}\{I_j^\gamma = 1\}$ that any subgraph of size k is a γ -quasi-clique are equal, whence the expected number of γ -quasi-cliques of size k in $G(n, p)$ is given by

$$\begin{aligned} \mathbb{E}[N_k^\gamma] &= \binom{n}{k} \mathbb{P}\{\text{a subgraph of size } k \text{ in } G \text{ is a } \gamma\text{-clique}\} \\ &= \binom{n}{k} \sum_{m=\lceil \gamma \binom{k}{2} \rceil}^{\binom{k}{2}} \binom{\binom{k}{2}}{m} p^m (1-p)^{\binom{k}{2}-m} \\ &= \binom{n}{k} \left[1 - \text{Bin} \left[\gamma \binom{k}{2}; \binom{k}{2}, p \right] \right], \end{aligned} \quad (8)$$

where $\text{Bin}(k; n, p)$ is the c.d.f. of the binomial distribution. As it will be seen, the integer $k = k_n^\gamma$ such that

$$\mathbb{E}[N_{k_n^\gamma}^\gamma] = 1 \quad (9)$$

for large values of n plays a central role in the sequel. The next proposition takes a first step in evaluating k_n^γ .

Proposition 1 *If $p < \gamma$, the integer $k = k_n^\gamma$ that satisfies $\mathbb{E}[N_{k_n^\gamma}^\gamma] = 1$ increases with n in such a way that $k_n^\gamma = o(n)$, $n \gg 1$.*

Proof. From expression (8) it is evident that $k = k_n^\gamma$ cannot be bounded for large values of n , since in that case the right-hand side of (8) would be equal asymptotically to $O(n^k)$. To verify that $k_n^\gamma = o(n)$, we construct an upper bound on the right hand side of equation (8). Using Stirling’s approximation, the binomial coefficient in (8) can be bounded as

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k. \quad (10)$$

To bound the summation term in (8), we use Chernoff’s bound for the tail of the binomial distribution [58]:

$$\sum_{i=m}^n \binom{n}{i} p^i (1-p)^{n-i} \leq \left(\frac{n-np}{n-m}\right)^{n-m} \left(\frac{np}{m}\right)^m,$$

where $m \geq np$. In our case $n = \binom{k}{2}$, $m = \lceil \gamma \binom{k}{2} \rceil$ (for simplicity,

we use $m = \gamma \binom{k}{2}$), and since $p < \gamma$, then $m \geq np$; thus, the requirement on m is valid. Thus,

$$\begin{aligned} & \sum_{i=\lceil \gamma \binom{k}{2} \rceil}^{\binom{k}{2}} \binom{\binom{k}{2}}{i} p^i (1-p)^{\binom{k}{2}-i} \\ & \leq \left(\frac{\binom{k}{2} - \binom{k}{2} p}{\binom{k}{2} - \gamma \binom{k}{2}} \right)^{\binom{k}{2} - \gamma \binom{k}{2}} \left(\frac{\binom{k}{2} p}{\gamma \binom{k}{2}} \right)^{\gamma \binom{k}{2}} \\ & = \left[\left(\frac{1-p}{1-\gamma} \right)^{1-\gamma} \left(\frac{p}{\gamma} \right)^\gamma \right]^k 2. \end{aligned} \quad (11)$$

Combining the upper bounds in (10) and (11), we have that if $k = k_n^\gamma$ satisfies $\mathbb{E}[N_k^\gamma] = 1$, whereby the following must hold for large enough values of n :

$$1 \leq \left(\frac{en}{k}\right)^k \left[\left(\frac{1-p}{1-\gamma} \right)^{1-\gamma} \left(\frac{p}{\gamma} \right)^\gamma \right]^k \binom{k}{2}. \quad (12)$$

Taking logarithm of the right hand side of the above inequality and dividing by k^2 , we obtain

$$\frac{1}{k} + \frac{\ln n}{k} - \frac{\ln n}{k} + \frac{1}{2} \left(1 - \frac{1}{k}\right) \ln c,$$

where the constant c has the form $c = \left(\frac{1-p}{1-\gamma}\right)^{1-\gamma} \left(\frac{p}{\gamma}\right)^\gamma$. It is easy to see from the inequality for arithmetic and geometric means that $c \in (0,1)$ for $0 < p < \gamma < 1$. Then, if k grows with n such that $\frac{k}{\ln n} = o(1)$, the above expression becomes negative for sufficiently large n , thereby contradicting the constructed upper bound (12). This implies that $k_n^\gamma = o(n)$, which proves the proposition.

Proposition 2 *If $p < \gamma$, the integer $k = k_n^\gamma$ that satisfies the equality $\mathbb{E}[N_{k_n^\gamma}^\gamma] = 1$, is given by*

$$k_n^\gamma = \frac{2}{\ln \left[\left(\frac{\gamma}{p}\right)^\gamma \left(\frac{1-\gamma}{1-p}\right)^{1-\gamma} \right]} \ln n + O(\ln \ln n), \quad n \gg 1. \quad (13)$$

In establishing Proposition 2 we rely on the following result due to [59].

Theorem 3 (McKay [59]). *Let $p \in (0,1)$ be fixed, and $p v \leq \kappa \leq v$ for some $v \geq 1$. Define $x = \frac{\kappa - p v}{\sigma}$, where $\sigma = \sqrt{v p (1-p)}$. Then*

$$\begin{aligned} & \sum_{i=\kappa}^v \binom{v}{i} p^i (1-p)^{v-i} = \\ & \sigma \binom{v-1}{\kappa-1} p^{\kappa-1} (1-p)^{v-\kappa} \frac{1-\Phi(x)}{\phi(x)} \exp \{(v, \kappa, p)/\sigma\}, \end{aligned} \quad (14)$$

where

$$0 \leq (v, \kappa, p) \leq \min \{ \sqrt{\pi/8}, x^{-1} \},$$

and $\Phi(x)$ and $\phi(x)$ are the cumulative and probability density functions of the standard normal distribution, respectively.

Proof of Proposition 2. Using the notations of Theorem 1, let

$$v = \binom{k}{2}, \quad \kappa = \gamma v, \quad \sigma = \sqrt{p(1-p)v}, \quad x = \frac{\gamma - p}{\sqrt{p(1-p)}} v^{1/2},$$

where we note that $\kappa \approx \lceil \gamma v \rceil > p v$ for large enough v , then the last term in (14) satisfies

$$\exp \{(v, \kappa, p)/\sigma\} = \exp \{O(v^{-1})\} = 1 + O(v^{-1}), \quad n \rightarrow \infty.$$

From the fact that x increases with n (cf. Proposition 1), it follows that

$$\frac{1-\Phi(x)}{\phi(x)} = \frac{1}{x} + O(x^{-3}) = \frac{\sqrt{p(1-p)}}{\gamma - p} v^{-1/2} (1 + O(v^{-1})), \quad n \rightarrow \infty,$$

where the well-known expansion

$$1 - \Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} (t^{-1} + O(t^{-3})), \quad t \rightarrow \infty,$$

was used. Invoking Stirling’s expansion for $\Gamma(z)$,

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}} z^z e^{-z} (1 + O(z^{-1})), \quad z \rightarrow \infty,$$

we obtain

$$\binom{v-1}{\gamma v-1} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\gamma}{1-\gamma}} [\gamma^\gamma (1-\gamma)^{1-\gamma}]^{-v} v^{-1/2} (1 + O(v^{-1})).$$

Thus, finally, the tail of the binomial distribution in (8) can be estimated as.

$$\sum_{i=\gamma v}^v v \binom{v}{i} p^i (1-p)^{v-i} = \frac{1}{\sqrt{2\pi}} \frac{1-p}{\gamma-p} \sqrt{\frac{\gamma}{1-\gamma}} v^{-1/2} \left[\binom{p}{\gamma} \left(\frac{1-p}{1-\gamma} \right)^{1-\gamma} \right]^v (1 + O(v^{-1})),$$

where $v = \binom{k}{2}$. Consequently, equation (8) can asymptotically be written as

$$\begin{aligned} & \binom{n}{k} \frac{1}{\sqrt{2\pi}} \frac{1-p}{\gamma-p} \sqrt{\frac{\gamma}{1-\gamma}} \binom{k}{2}^{-1/2} \\ & \left[\binom{p}{\gamma} \left(\frac{1-p}{1-\gamma} \right)^{1-\gamma} \right]^{\binom{k}{2}} \left(1 + O\left(\binom{k}{2}^{-1} \right) \right) = 1. \end{aligned} \quad (15)$$

Taking the logarithm of both sides of the last equality, we obtain

$$\begin{aligned} & -\left(k + \frac{1}{2}\right) \ln k + k + k \ln n + \ln c_1 - \frac{1}{2} \ln \binom{k}{2} + \binom{k}{2} \ln c_2 \\ & = O\left(\max\left\{\frac{k^2}{n}, \frac{1}{k}\right\}\right), \end{aligned} \quad (16)$$

where

$$c_1 = \frac{1}{2\pi} \frac{1-p}{\gamma-p} \sqrt{\frac{\gamma}{1-\gamma}}, \quad c_2 = \binom{p}{\gamma} \left(\frac{1-p}{1-\gamma} \right)^{1-\gamma}.$$

Taking into account that

$$\frac{1}{2} \ln \binom{k}{2} = \ln k - \ln \sqrt{2} + O(k^{-1}),$$

equation (16) can be written as

$$\begin{aligned} & \frac{k^2}{2} \ln c_2 - k \ln k + k(\ln n + 1 - \frac{1}{2} \ln c_2) - \frac{3}{2} \ln k + \ln \sqrt{2} c_1 \\ & = O\left(\max\left\{\frac{k^2}{n}, \frac{1}{k}\right\}\right). \end{aligned}$$

To obtain the main term of the asymptotical approximation of the solution of the last equation, let us restate it in the form.

$$\begin{aligned} & k^2 \left\{ \frac{1}{2} \ln c_2 + \frac{\ln n}{k} - \frac{\ln k}{k} + \frac{1 - \frac{1}{2} \ln c_2}{k} - \frac{3 \ln k}{2k^2} + \frac{\ln \sqrt{2} c_1}{k^2} \right\} \\ & = O\left(\max\left\{\frac{k^2}{n}, \frac{1}{k}\right\}\right). \end{aligned}$$

In view of the fact that $k = o(n)$ due to Proposition 1, the above expression can be further rewritten as

$$\frac{1}{2} \ln c_2 + \frac{\ln n}{k} + o\left(\frac{\ln n}{k}\right) = o(1),$$

whence we have that

$$k = \frac{2 \ln n}{\ln c_2^{-1}} + \chi(n), \quad \text{where } \chi(n) = o(\ln n).$$

To determine the order of the term $\chi(n)$, we restate the last equation as.

$$\frac{k}{2} \ln c_2 + \ln n - \ln k = O(1).$$

Writing down the expression for $k = k_n^\gamma$ in the form.

$$k_n^\gamma = \frac{2 \ln n}{\ln c_2^{-1}} + \chi(n),$$

and substituting it in the last equation, we obtain that $\chi(n) = O(\ln \ln n)$, which furnishes the statement of the proposition.

Next, we demonstrate that the number k_n^γ (given by Equation (13)), which solves the equation $\mathbb{E}[N_k^\gamma] = 1$, with probability 1 represents an upper bound on the size of the maximum γ -quasi-clique in a uniform random graph when $n \rightarrow \infty$. For this, we need the following property of γ -quasi-cliques.

Proposition 3 *If graph $G = (V, E)$, where $|V| = n$, is a γ -quasi-clique for some fixed $\gamma \in (0, 1]$, then for any $s < n$ there exists a γ -quasi-clique of size s in G .*

Proof. For $\gamma = 1$, this property is trivial. In the case of $\gamma \in (0, 1)$, it suffices to show that the statement of the proposition holds for $s = n - 1$. Since $G = (V, E)$ is a γ -quasi-clique, then

$$|E| \geq \gamma \binom{n}{2}.$$

Assume that there exists a vertex $i \in V$ with $\deg_G(i) \leq \gamma(n-1)$. Let $V_i = V \setminus i$; then the induced subgraph $G[V_i] = (V_i, E_i)$ is also a γ -quasi-clique, since

$$|E_i| \geq \gamma \binom{n}{2} - \gamma(n-1) = \gamma \binom{n-1}{2}.$$

If there is no such a vertex, i.e., $\deg_G(i) > \gamma(n-1)$ for all $i \in V$, then let $j = \operatorname{argmin}_{i \in V} \deg_G(i)$ be the vertex of G with the smallest degree, $\deg_G(j) = m > \gamma(n-1)$. As before, denote $V_j = V \setminus j$, and observe that the cardinality of the set of edges E_j of the induced subgraph $G[V_j] = (V_j, E_j)$ satisfies

$$\begin{aligned} |E_j| &= \frac{1}{2} \sum_{i \in V_j: (i,j) \in E} \deg_{G[V_j]}(i) + \frac{1}{2} \sum_{i \in V_j: (i,j) \notin E} \deg_{G[V_j]}(i) \\ &\geq \frac{1}{2} m(m-1) + \frac{1}{2} (n-m-1)m \\ &= \frac{1}{2} (n-2)m > \gamma \binom{n-1}{2}, \end{aligned}$$

thus verifying the statement of the proposition.

Proposition 4 Let M_n^γ denote the (random) size of the maximum γ -quasi-clique in a uniform random graph $G(n, p)$. If $0 < p < \gamma \leq 1$, then

$$\limsup_{n \rightarrow \infty} \frac{M_n^\gamma}{\ln n} \leq \frac{2}{\ln 1/p^*} \quad a.s., \quad (17)$$

$$\text{where } p^* = \left(\frac{p}{\gamma}\right)^\gamma \left(\frac{1-p}{1-\gamma}\right)^{1-\gamma}.$$

Proof. First, observe that

$$\mathbb{P}\{M_n^\gamma \geq k\} = \mathbb{P}\{N_k^\gamma \geq 1\} \leq \mathbb{E}[N_k^\gamma], \quad (18)$$

where the equality is due to Proposition 3. Define a sequence $k_n = \frac{2}{\ln 1/p^*} \ln n$, $n \geq 1$; then, from expression (8) for $\mathbb{E}[N_k^\gamma]$, one obtains by following the steps in Proposition 2 that for sufficiently large values of n

$$\mathbb{P}\{M_n^\gamma \geq k_n\} \leq \binom{n}{k_n} (p^*)^{\binom{k_n}{2}} \leq \frac{n^{k_n}}{\Gamma(k_n+1)} (p^*)^{\binom{k_n}{2}}. \quad (19)$$

From the definition of k_n it follows that the term

$$n^{k_n-1} (p^*)^{\binom{k_n}{2}}$$

is bounded for large enough n , whence the sought probability can be subsequently bounded as

$$\mathbb{P}\{M_n^\gamma \geq k_n\} \leq \frac{n}{\Gamma(k_n+1)} \leq \frac{ne^{k_n}}{k_n^{k_n}}.$$

Again recalling the definition of k_n , we note that

$$n^2 \frac{ne^{k_n}}{k_n^{k_n}} = \frac{\exp\{k_n(1 + \frac{3}{2} \ln 1/p^*)\}}{k_n^{k_n}} \rightarrow 0, \quad n \rightarrow \infty, \quad (20)$$

which implies that $\mathbb{P}\{M_n^\gamma \geq k_n\} = o(n^{-2})$ for $n \gg 1$, whereby the upper bound (17) on the size of the maximum γ -quasi-clique holds almost surely by virtue of the Borel-Cantelli lemma.

The next corollary shows that in sufficiently large random graphs $G(n, p)$, the size of the maximum γ -clique is almost surely above a certain value of the order of $\ln n$. It uses a well known fact, established by [25], that the size of the maximum clique in a uniform random graph $G(n, p)$ converges almost surely to $2 \ln n / \ln p^{-1}$. Note that M_n^1 represents the size of the maximum clique ($\gamma=1$) in a uniform random graph $G(n, p)$. Observe also that, according to (13),

$$\lim_{\gamma \rightarrow 1-0} k_n^\gamma = \frac{2 \ln n}{\ln 1/p} + O(\ln \ln n),$$

which corresponds to the well-known expression for the size of the maximum clique in uniform random graphs [24,25]. This allows us to define k_n^1 as the limiting value of k_n^γ above.

Corollary 1 If $0 < p < \gamma \leq 1$, then the size M_n^γ of the maximum γ -quasi-clique in a uniform random graph $G(n, p)$ satisfies

$$\begin{aligned} \frac{2}{\ln 1/p} &\leq \liminf_{n \rightarrow \infty} \frac{M_n^\gamma}{\ln n} \leq \limsup_{n \rightarrow \infty} \frac{M_n^\gamma}{\ln n} \leq \\ &\quad \frac{2}{\ln \left[\left(\frac{p}{\gamma}\right)^\gamma \left(\frac{1-p}{1-\gamma}\right)^{1-\gamma} \right]} \quad a.s. \end{aligned} \quad (21)$$

where k_n^γ is given by (13).

Proof. This follows immediately from Proposition 4, and the observation that, for any $\gamma < 1$, the size of the maximum γ -clique in $G(n, p)$ is always greater than the size of the maximum clique in the same graph, i.e.,

$$\mathbb{P}\{M_n^1 \leq M_n^\gamma\} = 1, \quad \gamma < 1.$$

Also note that the relations $0 < p < \gamma < 1$ imply that

$$\left(\frac{\gamma(1-p)}{p(1-\gamma)}\right)^{1-\gamma} > 1 > \gamma,$$

from which one infers the inequality

$$\binom{p}{\gamma}^\gamma \binom{1-p}{1-\gamma}^{1-\gamma} > p,$$

verifying that inequality for the lower and upper bounds on $M'_n/\ln n$ in (21) always holds, given the above assumptions on the values of p and γ .

Remark 5 From Fatou’s lemma it follows that bounds (21) on the maximum γ -quasi-clique size M'_n , which hold with probability 1, also hold for the mean maximum γ -quasi-clique size, $\mathbb{E}[M'_n]$:

$$\begin{aligned} \frac{2}{\ln 1/p} &\leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[M'_n]}{\ln n} \leq \limsup_{n \rightarrow \infty} \frac{\mathbb{E}[M'_n]}{\ln n} \\ &\leq \ln \left[\left(\frac{\gamma}{p} \right)^\gamma \left(\frac{1-\gamma}{1-p} \right)^{1-\gamma} \right]. \end{aligned} \quad (22)$$

In such a way, we have established that for any fixed $p < \gamma$ the asymptotic size of the maximum γ -quasi-clique is of the order of $\ln n$. Intuitively, when $p \geq \gamma$, the entire graph $G(n, p)$ becomes a γ -quasi-clique, thus the size of the maximum γ -quasi-clique has the order of n . Therefore, the natural question arising here is what happens when γ is fixed and p approaches γ . We show that there is a first-order phase transition in the asymptotic behavior of the order of magnitude of the maximum γ -quasi-clique in the point $p = \gamma$.

Proposition 5 If M'_n is the size of the maximum γ -quasi-clique in a uniform random graph $G(n, p)$ for some fixed $\gamma \in (0, 1)$, then with high probability (w.h.p.)

$$\lim_{p \nearrow \gamma} \lim_{n \rightarrow \infty} \frac{M'_n}{n} = 0, \quad (23a)$$

but

$$\lim_{p \searrow \gamma} \lim_{n \rightarrow \infty} \frac{M'_n}{n} = 1. \quad (23b)$$

Proof. The first limiting case follows from Proposition 4, since we proved that for any fixed $\gamma > p$ with probability 1

$$\lim_{n \rightarrow \infty} \frac{M'_n}{n} = 0.$$

To prove the equality in (23b), let X_{ij} be a Bernoulli random variable which is equal to 1 if there exists an edge (i, j) in the uniform random graph $G(n, p)$. Then,

$$\mathbb{P}\{M'_n = n\} = \mathbb{P}\left\{\sum_{(i,j)} X_{ij} \geq \gamma \binom{n}{2}\right\} = \mathbb{P}\left\{\frac{\sum_{(i,j)} X_{ij}}{\binom{n}{2}} \geq \gamma\right\}.$$

From the weak law of large numbers it follows that for any fixed $\varepsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{\sum_{(i,j)} X_{ij}}{\binom{n}{2}} - p\right| \leq \varepsilon\right\} \rightarrow 1, \quad n \rightarrow \infty.$$

Letting $\varepsilon = p - \gamma$, we obtain

$$\begin{aligned} \mathbb{P}\left\{\frac{\sum_{(i,j)} X_{ij}}{\binom{n}{2}} \geq \gamma\right\} &= \mathbb{P}\left\{p - \frac{\sum_{(i,j)} X_{ij}}{\binom{n}{2}} \leq \varepsilon\right\} \\ &\geq \mathbb{P}\left\{\left|p - \frac{\sum_{(i,j)} X_{ij}}{\binom{n}{2}}\right| \leq \varepsilon\right\} \rightarrow 1, \quad n \rightarrow \infty. \end{aligned}$$

Therefore,

$$\mathbb{P}\{M'_n = n\} \rightarrow 1, \quad n \rightarrow \infty,$$

which ends the proof of the proposition.

Linear Mixed-Integer Formulation of the Maximum γ -Quasi-Clique Problem

In this subsection we summarize a linear mixed-integer formulation of the maximum γ -quasi-clique problem [41] that was used for “small-scale” computational experiments as a part of the Computational Experiments section below.

Consider a graph $G = (V, E)$ with n vertices and an adjacency matrix A (with elements $a_{ij} = 1$ if there is an edge between vertices i and j , and $a_{ij} = 0$ otherwise), and suppose that one selects some subgraph G_s of G . In order to verify whether G_s is a γ -quasi-clique, we use the binary vector of variables $x \in \{0, 1\}^n$, where $x_i = 1$ if vertex i belongs to G_s , and $x_i = 0$ otherwise. The subgraph G_s is a γ -quasi-clique if the cardinality of its set of edges is at least

$$\begin{aligned} \gamma \binom{|G_s|}{2} &= \frac{1}{2} \gamma \sum_{i=1}^n x_i \left(\sum_{j=1}^n x_j - 1 \right) = \frac{1}{2} \gamma \left(\sum_{i,j=1}^n x_j x_i - \sum_{i=1}^n x_i \right) \\ &= \frac{1}{2} \gamma \left(\sum_{\substack{i,j=1 \\ i \neq j}}^n x_j x_i + \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \right) \\ &= \frac{1}{2} \gamma \sum_{\substack{i,j=1 \\ i \neq j}}^n x_j x_i, \end{aligned}$$

where the last equality is due to $x_i^2 = x_i$. The number of edges in the subgraph G_s can be calculated as

$$\frac{1}{2}x^T Ax = \sum_{\substack{i,j=1 \\ i \neq j}}^n a_{ij}x_jx_i.$$

Therefore, the problem of finding the maximum γ -quasi-clique in the graph G can be formulated as follows:

$$\begin{aligned} \max_{x \in \{0,1\}^n} \quad & \sum_{i=1}^n x_i \\ \text{s. t.} \quad & \sum_{\substack{i,j=1 \\ i \neq j}}^n a_{ij}x_jx_i \geq \gamma \sum_{\substack{i,j=1 \\ i \neq j}}^n x_jx_i. \end{aligned} \quad (24)$$

This is a 0–1 integer programming (IP) problem with a linear objective and a nonconvex quadratic constraint. A linearization of this problem can be performed at the expense of introducing additional variables and constraints.

A mixed-integer linear formulation of (24) with $O(n)$ variables and constraints is given below. Recall that originally we had only one constraint,

$$\sum_{\substack{i,j=1 \\ i \neq j}}^n (a_{ij} - \gamma)x_jx_i \geq 0,$$

which can be rewritten as

$$\sum_{i=1}^n x_i \left(\gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j \right) \geq 0.$$

Define the variables $w_i \in \mathbb{R}$, $i = 1, \dots, n$, as follows

$$w_i = x_i \left(\gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j \right), \quad i = 1, \dots, n.$$

Next, observe that each of the quadratic equalities above is equivalent to four linear inequalities

$$w_i \leq nx_i,$$

$$w_i \geq -nx_i,$$

$$w_i \geq \gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j - (1 - x_i)n,$$

$$w_i \leq \gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j + (1 - x_i)n,$$

$$x_i \in \{0,1\}, w_i \in \mathbb{R}.$$

Therefore, the problem of finding a maximum γ -quasi-clique can be represented as the following mixed integer linear programming problem with $2n$ variables (n binary variables and n continuous variables) and $4n + 1$ constraints. This formulation will be used for the computational experiments below.

$$\begin{aligned} \max \quad & \sum_{i=1}^n x_i \\ \text{s. t.} \quad & \sum_{i=1}^n w_i \geq 0, \\ & w_i \leq nx_i, \quad i = 1, \dots, n, \\ & w_i \geq -nx_i, \quad i = 1, \dots, n, \\ & w_i \geq \gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j - (1 - x_i)n, \quad i = 1, \dots, n, \\ & w_i \leq \gamma x_i + \sum_{j=1}^n (a_{ij} - \gamma)x_j + (1 - x_i)n, \quad i = 1, \dots, n, \\ & x_i \in \{0,1\}, w_i \in \mathbb{R}, \quad i = 1, \dots, n. \end{aligned} \quad (25)$$

Computational Experiments

In this subsection we consider exact and heuristic numerical computations of the size of the maximum γ -quasi-clique in randomly generated graphs. In particular, these computational experiments are aimed at checking the following two aspects: (i) how realistic the asymptotic bounds (21), (22) on the size of the maximum γ -quasi-clique M_n^γ and its mean value are for relatively small values of n ($n \sim 10^2$), and (ii) whether the approximate behavior of the relative size of the maximum γ -quasi-clique in *moderate-size* ($n \sim 10^4$) random graphs (as $p \nearrow \gamma$) exhibits the “step function” pattern, which was proven for $n \rightarrow \infty$ in Theorem 2.

According to Remark 5, in *large enough* random graphs $G(n, p)$ the average size $\mathbb{E}[M_n^\gamma]$ of the maximum γ -quasi-clique belongs to the interval

$$[m_n^1, m_n^\gamma] = \left[\frac{2 \ln n}{\ln\left(\frac{1}{p}\right)}, \frac{2 \ln n}{\ln\left(\frac{\gamma}{p}\right)^\gamma \left(\frac{1-\gamma}{1-p}\right)^{1-\gamma}} \right], \quad (26)$$

provided that $p < \gamma$. Therefore, it was of interest to check the applicability of the above bounds for relatively small values of n .

To this end, in the first set of computational experiments we generated a number of instances of uniform random graphs $G(n, p)$ with $n = 100$ and p ranging from 0.05 to 0.15; namely, we generated 100 instances of $G(100, p)$ for every p . Then, we employed the MIP formulation (25) to find the maximum γ -cliques in the generated graphs for values $\gamma = 0.9$ and $\gamma = 0.85$. Such a choice of parameters is justified by relatively better numerical tractability of the MIP problem (25) for sparse graphs. We used FICO™ Xpress Optimization Suite 7.1 [60] to solve the resulting instances of problem (25). The resulting average values of M_{100}^γ , as well as the minimum and maximum values of M_{100}^γ over 100 instances for each p , are reported in Table 1.

In the second set of computational experiments, we analyzed the behavior of the relative size of the maximum γ -quasi-clique for a fixed γ and different values of p and n . For $\gamma=0.7$, two sequences of values of parameter p were defined: $\{p_1, \dots, p_{100}\} = \{0.006, 0.012, \dots, 0.600\}$ and $\{p_{101}, \dots, p_{200}\} = \{0.601, 0.602, \dots, 0.700\}$. Note the smaller “step size” of the second sequence, which allows for a more thorough investigation of the maximum γ -clique size when the value of p becomes close to γ . The same experiment setup (with an appropriate adjustment of the sequence of values of p) was used for $\gamma=0.5$. For each value of $n=500, 1,000, 5,000, 10,000, 20,000$ and p_i from the sequence defined above, we generated instances of uniform random graphs $G(n, p_i)$, and determined the maximum γ -quasi-clique size M'_n . Since the MIP formulation (25) becomes computationally intractable for large dense graphs, we used maximum γ -quasi-clique GRASP heuristics due to [14], which were reported to perform quite well in massive graphs. Figures 2 and 3 report the results of the described computational experiments.

References

- Achlioptas D, D'Souza RM, Spencer J (2009) Explosive percolation in random networks. *Science* 323: 1453–1455.
- Ziff RM (2009) Explosive growth in biased dynamic percolation on two-dimensional regular lattice networks. *Phys Rev Lett* 103: 045701.
- Radicchi F, Fortunato S (2009) Explosive percolation in scale-free networks. *Phys Rev Lett* 103: 168701.
- Araújo NAM, Herrmann HJ (2010) Explosive percolation via control of the largest cluster. *Phys Rev Lett* 105: 035701.
- Radicchi F, Fortunato S (2010) Explosive percolation: A numerical analysis. *Phys Rev E* 81: 036110.
- Nagler J, Levina A, Timme M (2011) Impact of single links in competitive percolation. *Nature Physics* 7: 265–270.
- Fan J, Liu M, Li L, Chen X (2012) Continuous percolation phase transitions of random networks under a generalized achlioptas process. *Phys Rev E* 85: 061110.
- da Costa RA, Dorogovtsev SN, Goltsev AV, Mendes JFF (2010) Explosive percolation transition is actually continuous. *Phys Rev Lett* 105: 255701.
- Grassberger P, Christensen C, Bizhani G, Son SW, Paczuski M (2011) Explosive percolation is continuous, but with unusual finite size behavior. *Phys Rev Lett* 106: 225701.
- Riordan O, Warnke L (2011) Explosive percolation is continuous. *Science* 333: 322–324.
- Riordan O, Warnke L (2012) Achlioptas process phase transitions are continuous. *Annals of Applied Probability* 22: 1450–1464.
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402: C47–C52.
- Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences* 100: 12123–12128.
- Abello J, Resende MGC, Sudarsky S (2002) Massive quasi-clique detection. In: *LATIN 2002: Theoretical Informatics*. Springer, Lecture Notes in Computer Science, 598–612.
- Abello J, Pardalos PM, Resende MGC (1999) On maximum clique problems in very large graphs. In: *In External Memory Algorithms*. American Mathematical Society, pp. 119–130.
- Janson S, Luczak MJ (2008) Asymptotic normality of the k -core in random graphs. *Annals of Applied Probability* 18: 1085–1137.
- Luczak T (1991) Size and connectivity of the k -core of a random graph. *Discrete Math* 91: 61–68.
- Pittel B, Spencer J, Wormald N (1996) Sudden emergence of a giant k -core in a random graph. *J Comb Theory Ser B* 67: 111–151.
- D'Souza RM (2009) Complex networks: Structure comes to random graphs. *Nature Physics* 5: 627–628.
- Brunato M, Hoos HH, Battiti R (2008) On effectively finding maximal quasi-cliques in graphs. In: *Maniczzo V, Battiti R, Watson JP, editors, Learning and Intelligent Optimization*, Springer-Verlag, pp. 41–55.
- Matsuda H, Ishihara T, Hashimoto A (1999) Classifying molecular sequences using a linkage graph with their pairwise similarities. *Theoretical Computer Science* 210: 305–325.
- Pei J, Jiang D, Zhang A (2005) On mining cross-graph quasi-cliques. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, KDD '05, pp. 228–238.
- Seidman SB, Foster BL (1978) A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology* 6: 139–154.
- Bollobás B, Erdős P (1976) Cliques in random graphs. *Mathematical Proceedings of Cambridge Philosophical Society* 80: 419–427.
- Grimmett GR, McDiarmid CJH (1976) On colouring random graphs. *Mathematical Proceedings of Cambridge Philosophical Society* 77: 313–324.
- Bhattacharyya M, Bandyopadhyay S (2009) Mining the largest quasi-clique in human protein interactome. In: *Proceedings of the 2009 International Conference on Adaptive and Intelligent Systems*. Washington, DC, USA: IEEE Computer Society, ICAIS '09, pp. 194–199.
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297: 1551–1555.
- Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.
- Bu D, Zhao Y, Cai L, Xue H, Zhu X, et al. (2003) Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research* 31: 2443–2450.
- Crenson MA (1978) Social networks and political processes in urban neighborhoods. *American Journal of Political Science* 22: 578–594.
- Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications*. Structural analysis in the social sciences. Cambridge University Press.
- Scott J (2000) *Social Network Analysis: A Handbook*. Sage Publications, 2nd edition.
- Davis S, Trapman P, Leirs H, Begon M, Heesterbeek J (2008) The abundance threshold for plague as a critical percolation phenomenon. *Nature* 454: 634–637.
- Boginski V, Butenko S, Pardalos PM (2005) Statistical analysis of financial networks. *Computational Statistics & Data Analysis* 48: 431–443.
- Boginski V, Butenko S, Pardalos P (2006) Mining market data: A network approach. *Computers & Operations Research* 33: 3171–3184.
- Huang WQ, Zhuang XT, Yao S (2009) A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications* 388: 2956–2964.
- Sim K, Li J, Gopalkrishnan V, Liu G (2006) Mining maximal quasi-bicliques to co-cluster stocks and financial ratios for value investment. In: *Proceedings of the Sixth International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, ICDM '06, pp. 1059–1063.
- Bailey-Kellogg C, Chainraj S, Pandurangan G (2005) A random graph approach to nmr sequential assignment. *Journal of Computational Biology* : 569–583.
- Krzakala F, Zdeborova L (2008) Potts glass on random graphs. *EPL* 81: 57005.
- Janson S (2007) On a random graph related to quantum theory. *Combinatorics, Probability & Computing* 16: 757–766.
- Pattillo J, Veremyev A, Butenko S, Boginski V (2013) On the maximum quasi-clique problem. *Discrete Appl Math* 161: 244–257.
- Alba RD (1973) A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology* 3: 113–126.
- Mokken RJ (1979) Cliques, clubs and clans. *Quality and Quantity* 13: 161–173.
- Wasserman S, Faust K (1994) *Social Network Analysis*. Cambridge University Press.
- Scott J (2007) *Social Network Analysis: A Handbook*. London: SAGE Publications, 2nd edition.
- Balasundaram B, Butenko S, Hicks IV (2011) Clique relaxations in social network analysis: The maximum k -plex problem. *Oper Res* 59: 133–142.
- Bourjolly JM, Laporte G, Pesant G (2002) An exact algorithm for the maximum k -club problem in an undirected graph. *European Journal of Operational Research* 138: 21–28.
- Balasundaram B, Butenko S, Trukhanov S (2005) Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization* 10: 23–39.

Recall that it was shown in above that with probability that approaches to 1 it holds that

$$\lim_{n \rightarrow \infty} \frac{M'_n}{n} = \begin{cases} 0, & p < \gamma, \\ 1, & p > \gamma, \end{cases}$$

or, in other words, the relative size of the maximum γ -quasi-clique in uniform random graphs as a function of the density p of the graph with high probability represents a step function as $n \rightarrow \infty$. The results presented in Figures 2 and 3 complement these theoretical findings and show the respective behavior for “medium-scale” random graph instances.

Author Contributions

Conceived and designed the experiments: VB DJ. Performed the experiments: AV. Analyzed the data: AV VB PK DJ. Wrote the paper: AV VB PK. Performed the theoretical analysis: AV VB PK.

49. Veremyev A, Boginski V (2012) Identifying large robust network clusters via new compact formulations of maximum k -club problems. *European Journal of Operational Research* 218: 316–326.
50. Veremyev A, Boginski V (2012) Robustness and strong attack tolerance of low-diameter networks. In: Sorokin A, Murphey R, Thai MT, Pardalos PM, editors, *Dynamics of Information Systems: Mathematical Foundations*, Springer New York, volume 20 of Springer Proceedings in Mathematics & Statistics. pp. 137–156.
51. Erdős P (1947) Some remarks on the theory of graphs. *Bulletin of American Mathematical Society* 53: 292–294.
52. Chung F, Aiello W, Lu L (2001) A random graph model for power law graphs. *Experimental Mathematics* 10: 53–66.
53. Erdős P, Rényi A (1960) On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17–61.
54. Bollobás B (2001) *Random Graphs*. Cambridge University Press.
55. Matula DW (1970) On the complete subgraphs of a random graph. In: *Proceedings of the 2nd Conference on Combinatorial Mathematics and its Applications*. Chapel Hill: University of North Carolina, pp. 356–369.
56. Alon N, Spencer JH (2000) *The Probabilistic Method*. New York: John Wiley & Sons, Inc., 2nd edition.
57. Janson S, Luczak T, Rucinski A (2000) *Random Graphs*. New York: John Wiley & Sons, Inc.
58. Erdős P, Spencer JH (1974) *Probabilistic Methods in Combinatorics*. New York: Academic Press.
59. McKay BD (1989) On littlewood’s estimate for the binomial distribution. *Advances in Applied Probability* 21: 475–478.
60. (2010). FICO xpress optimization suite 7.1. <http://www.fico.com/en/Products/DMTools/Pages/FICO-Xpress-Optimization-Suite.aspx>.