# SCIENTIFIC REPORTS

**OPEN**

# Dynamic Networks that Drive the Process of Irreversible Step-Growth Polymerization

**Verena Schamboeck ⬤, Piet D. Iedema & Ivan Kryven ⬤**

**Many research fields, reaching from social networks and epidemiology to biology and physics, have experienced great advance from recent developments in random graphs and network theory. In this paper we propose a generic model of step-growth polymerisation as a promising application of the percolation on a directed random graph. This polymerisation process is used to manufacture a broad range of polymeric materials, including: polyesters, polyurethanes, polyamides, and many others. We link features of step-growth polymerisation to the properties of the directed configuration model. In this way, we obtain new analytical expressions describing the polymeric microstructure and compare them to data from experiments and computer simulations. The molecular weight distribution is related to the sizes of connected components, gelation to the emergence of the giant component, and the molecular gyration radii to the Wiener index of these components. A model on this level of generality is instrumental in accelerating the design of new materials and optimizing their properties, as well as it provides a vital link between network science and experimentally observable physics of polymers.**

Within recent years, network theory became an indispensable tool in a broad range of applied sciences ranging from social psychology and epidemiology to transport engineering, biology and physics[1,2]. It allows us to study the spreading of rumours and ideas[3,4], but also the spreading of diseases within populations[5] and cascades of failures in the electricity grid[6]. These and alike works established the universal language of network science that is valid across disciplines: it is conceivable that a study on the neural network of the human brain may teach us to better optimise transportation networks in growing cities[7]. The beginnings of network science, by which we refer to a combination of graph and probability theories, are commonly linked with works of S. Milgram and P. Erdős, see for example ref.[8]. Nonetheless, not of least importance for the foundation of the field played works of Flory who proposed to use random graph-like structures to study hyperbranched and cross-linked polymers[9]. In fact, Flory might have been the first to show interest in the percolation phenomenon even before Erdős introduced his famous random graph concept[10]. Furthermore, the same scientists who applied percolation to polymerisation developed the percolation theory of networks by working extensively on the network and polymer sides of the problem[11–20]. Although network science developed as a separate vibrant discipline of its own ever since, current theories of hyperbranched polymers are largely based on the early developments of network science, and the modern viewpoint of complex networks is only starting to diffuse back into polymer chemistry where this theory has arguably originated[21–27].

Conventional polymer networks are formed by a process called polymerization, during which small molecules bind together by means of covalent bonds and form large molecular structures. The functionality of these molecules is typically limited by the underlying chemistry and the bonds appear symmetric or asymmetric depending on the nature of functional groups that the reactants of the binding reaction bear[28]. It is mainly due to the variations of their topologies that polymeric materials feature such a broad range of physical properties. One of the most common polymerisation processes is the step-growth polymerisation of multifunctional monomers. This process leads to hyperbranched polymers of disperse sizes and irregular topologies that undergo a phase transition in their connectivity structure during the course of the polymerisation[28]. This transition is closely related to the percolation on networks[3–6]. The phase transition is marked by emergence of the gel, that is the giant molecule that spans the whole volume[1,29]. Flory provided simple analytical expressions for the average molecular size in these systems and was first to explain the onset of gelation, but limited himself to monomers of prescribed type $A_n$ or $A_n + B_2$[9]. Later, Stockmayer presented a formal expression for the whole distribution of molecular

University of Amsterdam, Van't Hoff Institute for Molecular Sciences, Amsterdam, 1090 GE, The Netherlands. Correspondence and requests for materials should be addressed to V.S. (email: v.schamboeck@uva.nl)

sizes[30], however the practical use of this expression is limited due to combinatorial complexity of computations. Durand and Claude derived a more general analytical expression for averages of the molecular size distribution[31]. Considerable progress has been made for the case of multifunctional monomers of type $A_n$, which feature symmetric bonds[22,32,33], whereas among asymmetric multifunctional monomers only monomers of the type $AB_2$ have a known analytical expression for the molecular size distribution as was demonstrated by Ziff[11]. For these reasons, the search continued resulting in a wave of fast approximate methods: as in works of Kryven et al.[34,35], Wulkow et al.[36], Tobita[37], Hillegers and Slot[38]. Although these methods are computationally fast, the approximate methods are hard to adapt to new polymerisation schemes, and especially the schemes requiring description with multidimensional distributions. In the same time, multidimensional distributions naturally arise in master equations describing monomers with multiple functional groups.

Yet another approach that has been applied to polymer networks only recently, the Molecular Dynamics (MD) simulation, is especially attractive as it produces very detailed information on the structure of polymer networks[39,40]. Molecular dynamics simulations are notorious for being computationally expensive, and therefore, limited to small samples and short time scales. In our previous work[39] we have demonstrated on the case of an acrylate polymer featuring predominantly symmetric covalent bonds, that many of the MD-generated network properties can be also reproduced by the configuration model for undirected random networks[41,42]. Furthermore, the recent developments in directed configuration models[29,43] present an opportunity to develop a generic polymerisation framework that will cover asymmetrical bonds as well. The latter, despite posing a more complex mathematical problem, are also more ubiquitous in polymerisation chemistry, and especially in that of hyperbranched and super-molecular polymers[44–46].

The current paper presents a new look on exactly solvable expressions for hyperbranched polymers by utilising latest developments of random graph models[1,41,47]. Being inspired by the kinetic theory of Krapivsky, Redner, and Ben-Naim[17,48,49], we employ a two-stage approach: we first devise a kinetic model for the transformations the monomer units undergo in time, and then we construct a configuration random graph, which deduces the global properties of the network from the two-variate degree distribution that is obtained on the first stage. Analytical expressions are obtained for various distributional properties of the polymer resulting from step-growth polymerisation of random combination of arbitrary functional monomers. The advantages of the proposed random graph model are grounded in the generic applicability and analytical expressions that are also fast to compute.

## Results

A polymer is a large molecule that consists of many repeat units, the monomers, and is formed as a result of chemical reactions that lead to covalent bonding between the monomers. Step-growth polymerization does not require an initiator and occurs between monomers that carry reactive functional groups. Many polymers with real world applications are formed as a result of step-growth polymerisation. Figure 1 features a few important examples related to polyesters, polyamides, and polyurethanes. The maximum number of chemical bonds that a single monomer bears is limited by the number and type of functional groups that are present in this monomer. If a system consists of solely two-functional monomers, only linear polymers are formed in the course of the step-growth polymerisation. However, if some (or all) monomers have more than two functional groups, it is possible to form hyperbranched polymers and networks.

In many chemical systems, two monomers bind through an asymmetric reaction that occurs between functional groups of different kinds. When two functional groups of different kinds are reacting, for example, as in the reaction between an acid and an alcohol leading to an ester, we refer to one group as the A-group, and the other – as the B-group. This asymmetric reaction is at the main focus of the current paper. Symmetric reactions occurring between two groups of the same kind, e.g. two alcohols reacting to form an ether, have been covered elsewhere[22]. Figure 1 exemplifies this notation on a few cases of polymers that feature linear and branched topologies. For each case, we indicate the structural formulas of the relevant monomers, highlight the functional groups, and give the corresponding AB notation. In Figure 2 the asymmetric reaction between A and B groups is illustrated on the example of an $A_2$ monomer (a monomer with two A-groups) reacting with an $AB_2$ monomer (a monomer with one A-group and two B-groups).

Before introducing the random graph model, we briefly summarise the terminology commonly used in graph theory. A directed graph consists of nodes and directed edges connecting them. A subgraph of a graph, in which any two nodes are connected by an undirected path is called a weakly connected component. In this work, we drop the prefix weakly, and refer to these components as connected components. When representing a polymer system as a graph, the monomers are identified with nodes, the chemical bonds with edges and thus a polymer molecule with a whole connected component. As the two sides of a chemical bond in the AB-reaction are not identical, we represent this asymmetry with directed edges. Without loss of generality, the directionality is defined as pointing from the B-group towards the A-group. The graph representation and the mapping of a reacted A-/B-group to an in-/out-edge is depicted in Figure 2.

The number of bonds that a monomer is bearing equals to the degree of the corresponding node. Generally speaking, monomers with distinct numbers of bonds may have different concentrations. To capture these differences, we refer to the node degree distribution, $u(i, j)$, which defines the probability of a randomly chosen node to have $i$ adjacent in-edges, and $j$ adjacent out-edges, and therefore, $u(i, j)$ is proportional to the concentration monomers with $i$ and $j$ reacted A- and B-groups. Figure 3b, demonstrates that the directed degree distribution $u(i, j)$ and the projected undirected degree distribution $u(k) = \sum_{i+j=k} u(i, j)$ that ignores the direction of the edges may lead to different sizes for connected components. In this extreme example, the directed degree distribution only allows connected components of size $s = 3$, whereas the undirected degree distribution does not limit these sizes at all. It turns out that much of the global information about the polymer system can be deduced from the degree distribution.
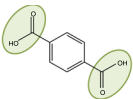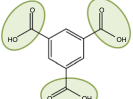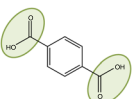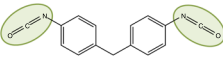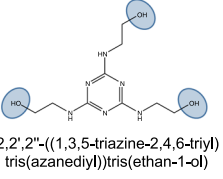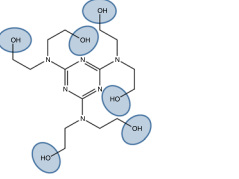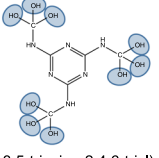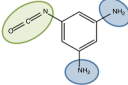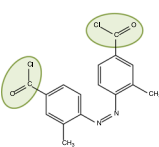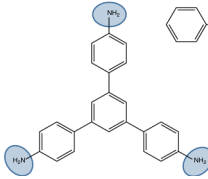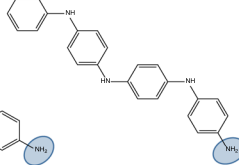
| Polymer class/ linear, branched A/B type | Monomers | AB representation |
|---|---|---|
| Polyester linear $A_2 + B_2$ | terephthalic acid          ethylene glycol |  |
| Polyester branched $A_3 + A_2 + B_2$ | benzene-1,3,5-tricarboxylic acid   terephthalic acid   ethylene glycol |  |
| Polyurethane branched $A_2 + B_3$ $+ B_6$ $+ B_9$ | Methylene diphenyl diisocyanate   2,2',2''-((1,3,5-triazine-2,4,6-triyl) tris(azanediyl))tris(ethan-1-ol)   2,2',2'',2''',2'''',2'''''-((1,3,5-triazine-2,4,6-triyl) tris(azanetriyl))hexakis(ethan-1-ol)   ((1,3,5-triazine-2,4,6-triyl) tris(azanediyl))tri(methanetriol) |  |
| Polyurea branched $AB_2$ | 5-isocyanatobenzene-1,3-diamine |  |
| Polyamide branched $A_2 + B_3 + B_1$ | (Z)-4,4'-(diazene-1,2-diyl) bis(3-methylbenzoyl chloride)   5'-(4-aminophenyl)- [1,1':3',1''-terphenyl]- 4,4''-diamine   N1-(4-aminophenyl)-N4- (4-(phenylamino)phenyl) benzene-1,4-diamine |  |

**Figure 1.** Examples of linear and branched polymers. The structural formulas of the monomers together with their AB representations that define the underlaying network topology are displayed. The list of polymers include: polyester[59], branched polyurethane[60], polyurea[61] and polyamide[62].

Hyperbranched molecules appear in a broad range of topologies that result from various chains of reactions, which occur between selected functional groups. One important feature here is the asymmetry of such reactions: they often occur between functional groups of complementary types (as opposed to symmetric reactions occurring between groups of identical types). To capture this variability we employ the configuration model that is defined by a directed degree distribution. The degree distribution reflects the state of the polymer system as driven by the chemical kinetics, and therefore, this distribution is time-dependent. The configuration model maximises entropy of all possible configurations that satisfy a given degree distribution at a time point of interest. This

Polymer representation
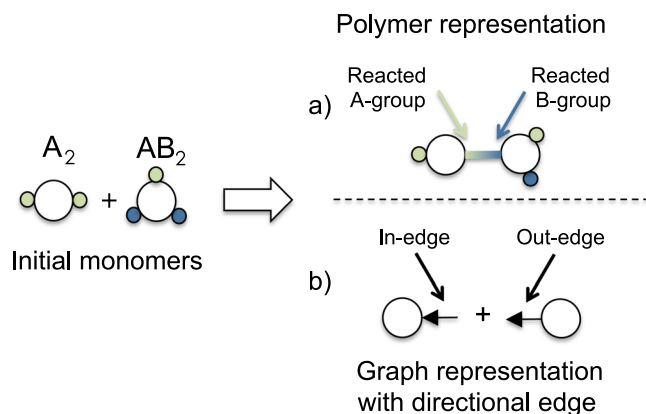


**Figure 2.** Illustration of an AB-reaction binding one A-group of an $A_2$ and one B-group of an $AB_2$ monomer. In (**a**) the colour of the bond indicates which monomer provided the A- and the B-group. In (**b**), the graph representation, the type of the functional group is stored in the directionality of the edge. An in-edge corresponds to a reacted A-group, an out-edge to a reacted B-group.
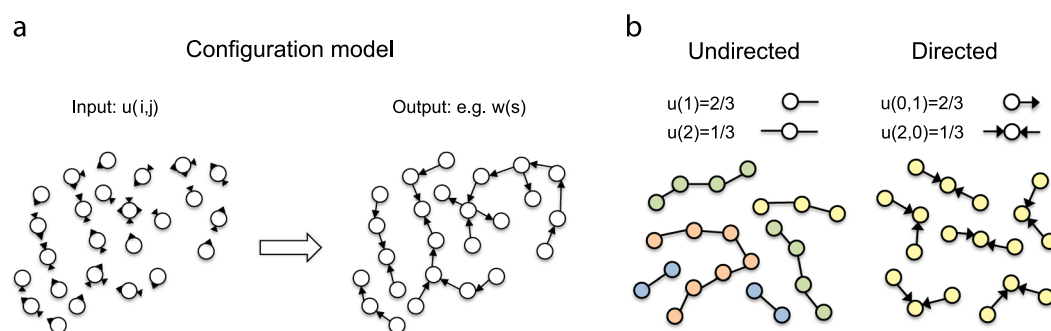


**Figure 3.** The concept of the directed configuration model. (**a**) (*left:*) Nodes are signed "half-edges" according to the bivariate degree distribution that provides an input for the model; (*right:*) these nodes are connected randomly so that the degree distribution is strictly satisfied. (**b**) An example of a simple degree distribution and the corresponding ensemble of connected components for the case of undirected and directed graphs. Note that in this example there is a qualitative difference between the structure of undirected and directed chains.

means that the model is egalitarian with respect to functional groups: every pair of complementary functional groups has equal probability to establish a bond. This principle is illustrated in Figure 3a.

The output of the configuration model is then processed to obtain various global properties of the polymer network, as for instance, the molecular weight distribution (probability that a randomly chosen node belongs to a weakly connected component of size $s$), the gel fraction (probability that a randomly chosen node belongs to the giant component), the mean-square gyration radii (related to the Wiener index of connected components), and the length of the average shortest path. The mathematical derivations of these results are presented in the Methods.

**Reaction kinetics and degree distribution.** The evolution of the degree distribution is governed by the reaction kinetics of the step-growth polymerization. This process converts an arbitrary pair of A- and B-groups into a bond, which is represented as an edge between nodes in the model. Since a monomer may carry multiple A- and B-groups, probability that a pair of monomers react is dependent on the number of unreacted functional groups they carry. The formal mechanism for the reaction between two monomers is given by:

$$(i, j, I, J) + (i', j', I', J') \xrightarrow{k_{AB}(I-i)(J'-j')} (i+1, j, I, J) + (i', j'+1, I', J'), \qquad (1)$$
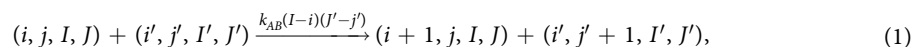
where vector $(i, j, I, J)$ denotes the state of a monomer: $I, J \geq 0$ are the numbers of respectively A- and B-groups on this monomer; $i = 0, 1, \ldots, I$ and $j = 0, 1, \ldots J$ denote the number of groups of respectively type A or B that have already been converted into bonds by the reaction. For each monomer, the indices $(I, J)$ are defined a priori, whereas $(i, j)$ change over time. The reaction rate is given by the product of the rate constant $k_{AB}$, the number of unreacted A-groups on the first monomer $(I - i)$, and the number of unreacted B-groups on the second monomer $(J' - j')$. The following assumptions are made: (1) the reactivity for any pair of A- and B-groups is equal; (2) the reactivity does not change throughout the process. Let $M_{i,j,I,J}(t)$ be the probability that a randomly chosen monomer has configuration $(i, j, I, J)$ at time $t$. This probability is proportional to the concentration of the monomers.

**Figure 4.** Comparison of the theory to data from experiments and computer simulations. (**a**) Experimental degree distribution as extracted form the $^{13}C$ NMR spectrum of hyperbranched polyester (HPE) in dimethylformamide (DMF)[50] compared to theoretical degree distribution in the $A_2$:$B_3$ = 1:1 system at $p_A = 0.93$. (**b**) Phenolic unit degree distribution (only out-edges) in a phenol-methylene system predicted by MD simulation (d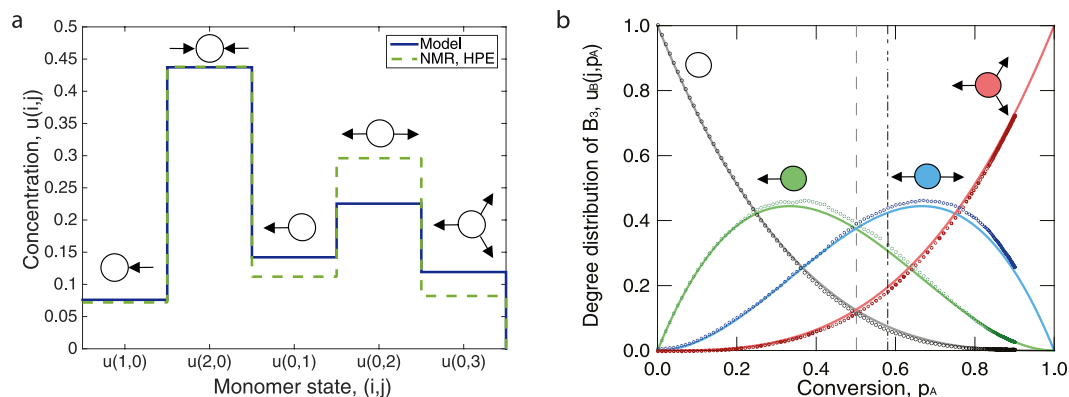oted lines)[40] compared to the theory (solid lines). The gel point is predicted at $p_A = 0.58$ by the MD simulation, the theory predicts it at $p_A = 0.5$.

The time variation $\frac{\partial M_{i,j,I,J}(t)}{\partial t}$ as governed by the process eq. (1) is described by the corresponding master equation, see Methods. In the general case of more than two distinct functional groups (e.g. A-, B-, C-, D-groups) and several reaction mechanisms (e.g. AB-, CD- reaction), the master equation becomes more complex due to a combinatorial number of monomer species of defined type and state. In that case, automated reaction networks can be applied to algorithmically construct the corresponding master equation[23].

The temporal degree distribution $u(i, j, t)$ is directly deduced from $M_{i,j,I,J}(t)$ by summating over all monomer types $I$, $J$. As discussed in Methods, the step-growth process eq. (1) leads to the following degree distribution $u(i, j, t)$:

$$u(i, j, t) = \sum_{I,J \geq 0} \binom{I}{i} p_A(t)^i (1 - p_A(t))^{I-i} \binom{J}{j} p_B(t)^j (1 - p_B(t))^{J-j} P(I, J).$$

(2)

This expression is given by the product of binomial distributions for the in- and out-edges, with $p_A(t)$ and $p_B(t)$ being the probability that a random A-/B-group is reacted (also referred to as A-/B-conversion), see eqs (12) and (13). The probability distribution $P(I, J)$ defines the initial concentration of monomer types and is referred to as the monomer functionality distribution. Probabilities $P(I, J)$ provide the sole input to the model.

In some cases, the degree distribution $u(i, j, t)$ can be measured directly by nuclear magnetic resonance (NMR). Due to the chemical shift in the NMR spectrum, every monomer state has its own distinct frequency. However, the bigger the monomer is and the more functional groups it has, the harder it is to identify the states. In Figure 4 we compare the degree distributions predicted by eq. (2) with the experimental NMR data from ref.[50], and also, for a different polymerisation system, against MD simulation data from ref.[40].

## Gelation point and gel fraction.

The gelation point marks the transition of the system from a liquid-like to a solid-like state during the polymerization process. After this transition, an increasing positive fraction of monomers becomes part of the single gel molecule. The transition is typically observed by measuring the fraction of the insoluble part of the polymer, or performing rheological experiments. From the network theory perspective, this is a well-known phenomenon as the gel transition point corresponds to the emergence of the giant component in the network topology. The size of the giant component is of the order of the whole system size, and we therefore quantify the gel size $g_f$ as the probability that a randomly chosen monomer belongs to the giant component. The Methods section explains how $g_f$ can be calculated if the degree distribution is known.

The above described process of polymerisation is closely related to percolation on networks. The phenomenon of percolation is always studied on a specific networks, e.g square lattice, Bethe lattice, or any arbitrary network. In polymer chemistry, it is mostly introduced as a process of randomly adding bonds to an empty template of the network[51]. This process is precisely reverse to the process of removing bonds from the same full network. Let $p$ be the probability that a randomly selected edge is not removed.

Under this notation, percolation can be thought of as a temporal process that starts at $p = 0$ and ends at $p = 1$. Moreover, this process is known to feature a phase transition at critical probability $p_{critical}$, the point at which the giant component appears. When there are equal amounts of A and B functional groups, this process is precisely reverse to the step-growth polymerisation where the edges are being added to a network randomly, and $p_{critical}$ coincides with the gel-point conversion of functional groups. If there are more B-groups than A-groups present in the system initially, the A-groups are limiting the reaction and we conventionally refer to the conversion of A groups as the conversion: $p = p_A$. Without loss of generality it may be assumed that the number of A-groups is less or equal to the number of B-groups. The moment in time when gelation occurs is completely defined by the proportion of monomers of different functionalities that are present initially in the system. Generally speaking, the higher the functionality, the earlier gelation occurs in time and/or conversion, and a precise quantitative estimate

| Directed | | Undirected | | Flory-Stockmayer | |
|---|---|---|---|---|---|
| $A_2B_2$ | $p_{\text{critical}} = \frac{1}{3}$ | $A_4$ | $p_{\text{critical}} = \frac{1}{3}$ | $A_4$ | $p_{\text{critical}} = \frac{1}{3}$ |
| $A_3{:}B_3 = 1{:}1$ | $p_{\text{critical}} = \frac{1}{2}$ | $A_3$ | $p_{\text{critical}} = \frac{1}{2}$ | $A_3$ | $p_{\text{critical}} = \frac{1}{2}$ |
| $AB_2{:}A_2B = 1{:}1$ | $p_{\text{critical}} = \frac{1}{2}$ | $A_3$ | $p_{\text{critical}} = \frac{1}{2}$ | $A_3$ | $p_{\text{critical}} = \frac{1}{2}$ |
| $A_2B_2{:}AB = 1{:}1$ | $p_{\text{critical}} = \frac{3}{7}$ | $A_4{:}A_2 = 1{:}1$ | $p_{\text{critical}} = \frac{3}{7}$ | — | — |
| $AB_2$ | $p_{\text{A,critical}} = 1$ | — | — | — | — |
| $A_2{:}B_3 = 1{:}3$ | no gelation* | — | — | — | — |
| $A_2{:}B_3{:}B_1 = 10{:}1{:}8$ | no gelation** | — | — | — | — |

**Table 1.** Examples of mappings between symmetric AB-systems and AA-systems and the corresponding critical conversion $p_{\text{critical}}$. The critical conversions in the first column are calculated by eq. (3), of the second by eq. (43)[22] and the third column by the Flory-Stockmayer criterium $\frac{1}{z-1}$. *Due to the depletion of A-groups. **Due to the significant concentration of terminal units $B_1$.
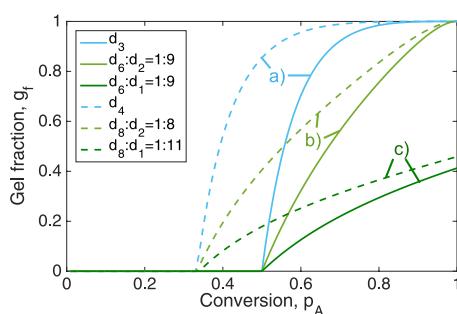


**Figure 5.** Illustration of three types of gel growth behaviour: (**a**) steep growth, (**b**) slow growth, (**c**) gel does not reach full system size. The different types of behaviour are cased by the composition of monomers. Two groups are depicted: (1) solid lines, gel point at $p_{\text{A,critical}} = 0.5$, (**a**) $A_3{:}B_3 = 1{:}1$, (**b**) $A_6{:}B_6{:}A_2{:}B_2 = 1{:}1{:}9{:}9$, (**c**) $A_6{:}B_6{:}A_1{:}B_1 = 1{:}1{:}9{:}9$; (2) dashed lines, gel point at $p_{\text{A,critical}} = 0.33$, (**a**) $A_4{:}B_4 = 1{:}1$, (**b**) $A_8{:}B_8{:}A_2{:}B_2 = 1{:}1{:}8{:}8$, (**c**) $A_8{:}B_8{:}A_1{:}B_1 = 1{:}1{:}11{:}11$.

of the gelation conversion is discussed in Methods. It turns out that one can connect the critical conversion directly to the functionality distribution $P(I, J)$:

$$p_{\text{critical}} = \frac{\nu_{01}}{\nu_{11} + \sqrt{(\nu_{02} - \nu_{01})(\nu_{20} - \nu_{10})}}. \tag{3}$$

where $\nu_{mn} = \sum_{I,J \geq 0} I^m J^n P(I, J)$. It is important to note that in the case of a symmetric functionality distribution $P(I, J) = P(J, I)$ (see also Methods), eq. (3) gives the same result as its counterpart for an AA-system (undirected system[22]) with $P(K) = \sum_{K=I+J} P(I, J)$. In the case of $P(K)$ being distinct from zero for only one $K$-functional species, eq. (3) simplifies to the Flory-Stockmayer equation for gelation $p_{\text{critical}} = \frac{1}{z-1}$, with $z = I + J$ the number of first neighbours. See Table 1 for illustration.

Equation (3) allows one to screen a vast number of systems and determine their gel-point conversions if such occur. We also can deduce from the theory discussed in Methods that some systems will never form gel. Here again, one can identify whether a system of given monomer functionalities gelates by studying $P(I, J)$. Namely, a monomer system forms gel if the following inequality is satisfied:

$$(\nu_{02} - \nu_{01})(\nu_{20} - \nu_{10}) - (\nu_{11} - \nu_{01})^2 > 0. \tag{4}$$

For the physical properties of the final material, both factors play a definitive role: when does gelation start and how does the gel fraction evolve in the course of the polymerisation? The growth rate of the gel fraction, $g_f$, is determined by the functionalities of the initial monomers and their concentration distribution. In Figure 5, a few examples illustrate different types of behaviour of the gel buildup. In these examples, we optimise the initial functionalities and concentrations of monomers to reach two final target properties: (1) a fixed gel point conversion of either $p_{\text{A,critical}} = 0.5$, as depicted by the solid lines, or $p_{\text{A,critical}} = 0.33$, as indicated by the dashed lines; (2) we distinguish three different types of growth behaviour: (a) a steep growth with most monomers being incorporated into the gel rapidly after gel point, (b) a slow growth with the gel reaching full size only at full conversion, (c) a slow growth with the gel never reaching the system size. Behaviour (a) is observed for systems with purely high-functional monomers, (b) for systems with few high-functional monomers and many 2-functional monomers, and (c) for few high-functional monomers and many 1-functional monomers that act as terminal units. The reason for the gel in (c) never reaching the full system size is the formation of small connected components that stop growing because of having all functional groups being capped with one-functional terminal units. For example, when a component is composed of one 6-functional monomer connected to six 1-functional monomers. The Methods section gives the general equations for the gel-point conversion and gel fractions.
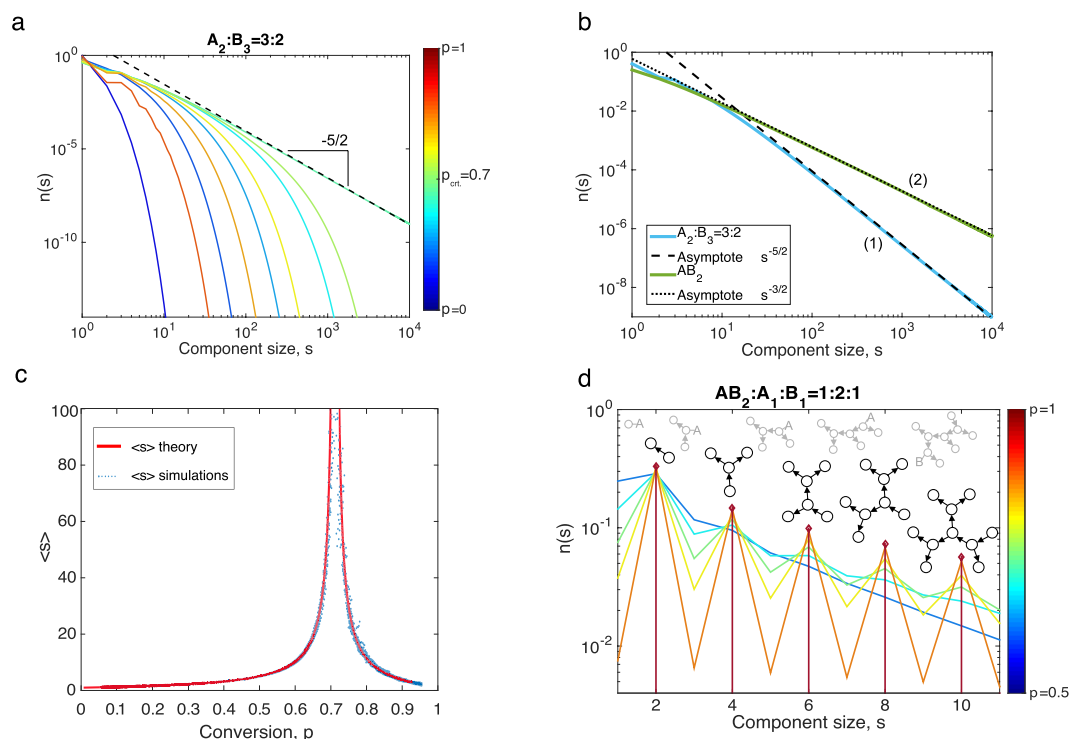
**Figure 6.** (**a**) Molecular size distributions of the system $A_2:B_3 = 3:2$ at different conversions as indicated by the colour scheme. At $p_{A,critical} = 0.707$ the distribution is scale-free, $n(s) \propto s^{-5/2}$. (**b**) Molecular size distributions $n(s)$ and their critical asymptotes as obtained for two systems: (1) $A_2:B_3 = 3:2$ and (2) $AB_2$. (**c**) Weight-average molecular weight for the system $A_2:B_3 = 3:2$ features singularity at $p_{critical} = 0.7$. (**d**) Oscillating size distributions of the system $AB_2:A_1:B_1 = 1:2:1$ at different conversions as indicated by the colour scheme. Examples of terminated (*black colour*) and living (*grey colour*) polymer topologies are indicated for reference.

These results are also interesting when studying polymer ageing and degradation. Consider a degradation process under which every chemical bond dissociates independently with equal probability. This process is reverse to the introduced polymerization process, and the gel fraction is a measure of how strongly the system is interconnected. Clearly, the systems of type (a) will show a different behaviour during degradation than systems of type (b). For (a), the system will stay connected for a long time, but will eventually collapse into many small pieces quite abruptly. Type (b) systems will show a more continuous, and therefore more predictable, degradation behaviour, which might be more desirable from application point of view.

**Molecular size distribution, averages and asymptotes.** From a network theory perspective, a separate polymer molecule is a connected component. The sizes of the latter are typically characterised by a size distribution. There are two common ways that such distributions can be defined. The molecular weight distribution $w(s)$ corresponds to the probability that a randomly chosen *monomer* belongs to a connected component of size $s$, whereas the molecular size distribution $n(s)$ is the probability that a randomly chosen *component* has size $s$. One can be converted into the other by an appropriate weighting and normalisation: $n(s) = Cs^{-1}w(s)$. In Methods we present an exact equation that connects the degree distribution with the molecular weight distribution $w(s)$. As a general rule, the exact values of $w(s)$ can be computed spending $\mathcal{O}(s^2 \log s)$ multiplicative operations, and in a special case of only one initial monomer type, the analytic expression for the molecular weight distribution is given in Methods.

The global behaviour that is observed in all polymerising systems can be summarised as follows: Initially, all monomers are unconnected, thus only molecules of size $s = 1$ are present. With increasing conversion $p_A$, larger molecules emerge. The size distribution features the exponential decrease at the tail, and becomes broader with progressing conversion until the gel point $p_{A,critical}$ is reached. Only at this single point the size distribution becomes scale-free. Figure 6a demonstrates this behaviour on an example. After the gel transition point, the size distribution describes only the soluble part of the system, so that the size of the gel is given by the gel fraction $g_f = 1 - \sum_{s=1}^{\infty} w(s)$. Furthermore, the size distribution returns to its exponential behaviour and becomes narrower with increasing conversion.

Surprisingly, there are two distinct types of polymerisation systems featuring different types of asymptotic behaviour. Most of polymer systems feature a size distribution with asymptote

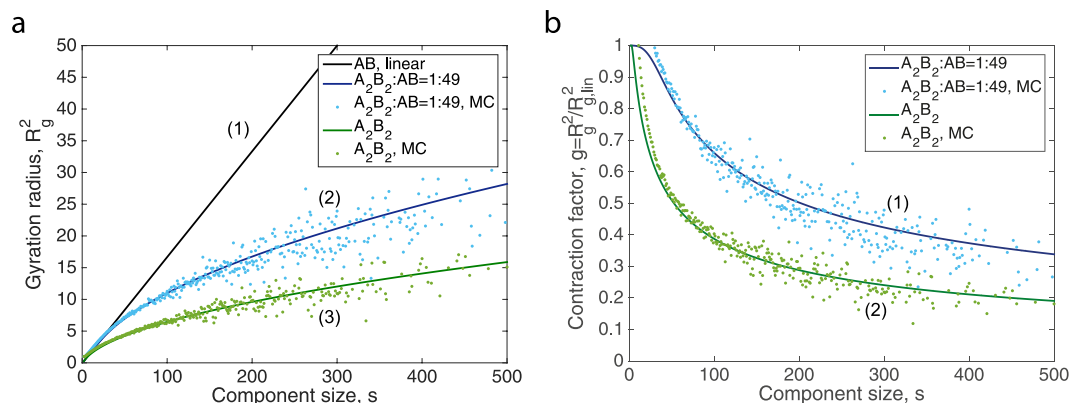$$n(s) \propto e^{-C_1 s} s^{-5/2}. \tag{5}$$

**Figure 7.** Comparison of the theory (*solid lines*) to simulation data (*scattered data*). The simulation data is the average over 100 generated networks that consist of $N = 10000$ nodes. (**a**) Gyration radius, three different systems are investigated: (1) linear AB, (2) sparsely branched $A_2B_2$:AB = 1:49, (3) hyperbranched $A_2B_2$. (**b**) Contraction factor of branched components, two systems are considered: (1) sparsely branched $A_2B_2$:AB = 1:49, (2) hyperbranched $A_2B_2$.

This, for instance, includes the $A_2$:$B_3 = 3$:2 system as illustrated in Figure 6a. However, some polymers may also feature a different asymptotic mode, namely

$$n(s) \propto e^{-C'_1 s} s^{-3/2}. \tag{6}$$

This asymptote arises in all systems of type $AB_n$, for $n > 1$. These two distinct types of asymptotic behaviours can also be attributed to the different types of kernel functions (additive versus product kernel) in the underlying aggregation equation[11,12]. In Figure 6b, this peculiar case of asymptotic behaviour is illustrated by comparing two very similar polymer systems $AB_2$ and $A_2 + B_3$ that yet feature different asymptotic modes.

The gel transition is also noticeable in the evolution of weight average molecular weight $M_w$, which features a singularity at the critical point. The evolution of $M_w$ as predicted by the theory is compared against stochastic simulations in Figure 6c. The figure shows good agreement between the theory and the numerical simulation except for critical conversion. At this point, the stochastic simulations (scatter plot) suffer form the small-system-size effect. The Methods section gives analytical equations for the weight average molecular weight in the pre-gel and gel regimes.

Interestingly, in some cases molecular size distributions feature oscillations. One such example is given in Figure 6d, depicting the system $AB_2$:$A_1$:$B_1 = 1$:2:1. At full conversion $p_A = 1$ (dark red line) only molecules of specific favoured sizes are present. At lower conversions, also other sizes occur that exhibit strongly reduced probability as compared to the favoured sizes. In this example the concentration of odd-size component is favoured, however, depending on the distribution of functionalities such oscillations may occur with arbitrary large "periods". It turns out, that the monomers of functionality one play an important role in these oscillations, as they terminate the growth of polymer molecules and thus fix sizes of these molecules at a constant value.

**Gyration radius.**    Consider a branched polymer molecule that is composed of $s$ monomers. The actual volume this molecule spans is related to how 'branched' this molecule is. In systems that contain no gel, or are below the gel transition, it is conventional to characterise this volume by the a quantity called gyration radius $R_g(s)$, which can also be estimated by light scattering experiments in a polymer solution[52].

Linear chains feature $R_{g,\text{lin}}(s) = b\sqrt{\frac{s}{6}}$, where $b$ is the Kuhn length. In Methods, we derive the analytical equation that links the degree distribution and the mean square gyration radius for $s \gg 1$. Figure 7a shows how one can influence $R_g(s)$ by tuning the set of initial monomers. This figure also compares the theoretical gyration radii against gyration radii obtained form stochastically generated networks. An alternative way of looking at the gyration radius is the contraction factor[53], which is defined by $g(s) = R_g^2(s)/R_{g,\text{lin}}^2(s)$. The contraction factor tells us how much more compact the actual molecule is in comparison to the linear one having the same number of monomers. Figure 7b compares the theoretical predictions of this quantity versus simulations.

Another unexpected result that is revealed by directed random graph theory, is that during the progress of step-growth polymerisation, for all molecules of size $s$ the mean gyration radius is constant in time. Since the size distribution does change in time, this time-independency is lost if one calculates an average of this quantity over different molecular sizes. It is likely that other than step-growth polymerisation processes do not feature time-invariant gyration radii.

The gyration radius is also directly proportional to the Wiener index[54], which is another topological index to characterise branched molecules. The Wiener index $\mathcal{W}(s)$ is defined as the sum of the lengths of the shortest paths between all pairs of the monomer units in the molecule. The relation between the mean square gyration radius and the Wiener index is given by $\mathcal{W}(s) = \frac{R_g^2(s)}{s^2}$.

**Scaling of the node neighbourhood size and criticism of well-mixing assumption.** Since the radius of gyration characterises only finite-sized molecules rather than the gel, which is virtually infinite in size, we are in need of devising an additional measure for the degree of connectedness of the gel. With this aim, we investigate the number of nodes at distance $l$ from a randomly selected one, which we shortly refer to as $\mathcal{B}_l$. Here, we utilise the topological notion of distance: nodes $i$ and $j$ are at distance $l$ if the shortest undirected path connecting these nodes has length $l$. Since the giant component is infinite in size, the larger the distance $l$ the more nodes are incorporated in the volume of a sphere. In fact, we are mainly interested in the way this quantity asymptotically depends on $l \gg 1$. Newman derived an expression for a similar quantity for directed paths[41], whereas in the polymer context we are interested in the weak sense of connectivity. This means that we consider the shortest paths that does not respect the directionality of the bonds.

An analytical expression for this behaviour is derived in Methods. We observe that the average path length in the gel exponentially depends on $l$:

$$\mathcal{B}_l = Ce^{l/l_0}, \quad l \gg 1, \tag{7}$$

where $C$ and $l_0$ are constants. Note that structures that reside in three-dimensional Euclidean space must feature a different scaling law:

$$\mathcal{B}_l = CN^{d_f}, \tag{8}$$

where $1 \leq d_f \leq 3$ is the cluster growth dimension. The estimate given in eq. (7) is quite a discouraging result, as a network that features an exponential scaling cannot be physically embedded in the Euclidean space under the condition that the nodes are uniformly distributed with constant density. This unphysical exponential growth of a node neighbourhood is caused by one of the fundamental assumptions made in the random graph model, and many other popular models that do not explicitly track the spatial configuration of the network also suffer from the same criticism. In fact, we refer here to one of the most commonly used assumptions of a well-mixed system: any two functional groups react with an equal probability irrespectively of their location in the topology. In real systems, however, monomers interact with the rest of the network that may locally hinder them to react. That said, it is important to note, that the scaling given by eq. (7) does not hold before, and precisely at, the gel transition, and the well-mixed system assumption may remain a good approximation in these regimes.

## Discussion

This paper employs recent developments in network science to formalise the step-growth polymerisation process as a problem fully described by network generation. In order to do so, we proposed to view the polymer architecture as a directed network that is described by a dynamic degree distribution. Although the physical connection between monomers by means of covalent bonding is completely symmetrical, the directionality of the edges keeps record of the asymmetry of the chemical reaction that created the corresponding covalent bond. This approach allows a classification and a general treatment of a vast range of real-world polymerisation problems and can be used for optimisation and design of new materials. As a general rule, the parameters of step-growth polymerising systems comprise a high-dimensional parameter space that dictates the reaction kinetics, network structures, and physical properties of the final material. Therefore, it is important to have a fast way to map the polymerisation parameters to the final topological properties of the polymer network.

We have matched various idioms present in polymer chemistry to corresponding graph-theoretical analogues and indicated how these can be predicted knowing the input parameters of the system by means of analytical expressions. For instance, we gave analytical quantifications of the gelation time, the topological phase transition and the associated to it molecular weight singularity, the molecular size distribution and its asymptotes, the gyration radii of polymers, and the scaling of the monomer neighbourhood size. Some of these findings also provide an unexpected qualitative insight on chemistry of polymerisation. For instance, we have revealed the existence of two asymptotical modes that appear in the molecular size distribution, the fact that these size distributions might feature a peculiar oscillating behaviour and that the gyration radii in step-growth polymerised molecules are not dependent on time but only on the sizes of these molecules.

On a broader scope, this work paves the way to viewing polymer networks, as well as other types of network formations in condensed matter, as being complex networks in which the topology and additional layers of information (multiplexity) all contribute to the formation and function of the macroscopic material. The theoretical framework that addresses multiplexity is a topic of current network science, and one of the goals of the authors is to advance the understanding of multiplex networks in condensed matter in their future works.

Although they were produced to aid polymer chemistry in the first place, these findings are also relevant to a broader network science community as one can view polymerisation as a process that is related to percolation. In this way, the paper amounts to understanding percolation in directed networks, which turn out to feature a richer behaviour then in undirected networks. Moreover, maintaining the link to polymers and polymerisation allows one to validate complex network theories with physics. Newly developed theories of such kind have been predominantly compared to experimental data derived from single point observations, which are hard to reproduce. For example, we cannot grow the Internet from scratch again. Yet, the degree distribution may be measured by NMR and more experimental techniques that will extract network properties from matter are on their way.

We finalise the paper with a note of caution that is addressed to the whole modelling community of polymer networks. The network analysis of the node-neighbourhood scaling in the configuration model points out the existence of unphysical features that appear after the gel transition. Importantly, the unphysical scaling is not an artefact of our approach but rather an implication of the commonly trusted assumption of chemical systems being well-mixed. This assumption is standard for many modelling methods that do not account for spacial embedding of the chemical

species in the three-dimensional space, as for instance is the case for the rate equations, Flory-Stockmayer theory, population balance equations, kinetic Monte Carlo, and other methods. We therefore would like to encourage the search of new network models that bring together mutual interaction of the topology and space.

## Methods

**Master equation of the degree distribution.** In this subsection we briefly summarise the theory from ref.[29] that allows us to recover the time evolution of the bivariate degree distribution by constructing an analytically solvable master equation. We distinguish monomer species by counting the numbers of functional groups of both types $I, J$ and the numbers of in- and out-edges $i, j$. During the progress of polymerisation the functional groups are converted into chemical bonds between the monomers, and the concentration profiles $M_{i,j,I,J}(t)$ obey the following master equation:

$$
\begin{aligned}
\frac{\partial}{\partial t} M_{i,j,I,J}(t) = {} & (I - i + 1)(\nu_{01} - \mu(t))M_{i-1,j,I,J}(t) \\
& + (J - j + 1)(\nu_{10} - \mu(t))M_{i,j-1,I,J}(t) \\
& - ((I - i)(\nu_{01} - \mu(t)) + (J - j)(\nu_{10} - \mu(t)))M_{i,j,I,J}(t).
\end{aligned}
\tag{9}
$$

Since initially, at $t = 0$, there are no bonds, the system is completely described by the distribution of functional groups: $M_{i,j,I,J}(0) = P(I, J)$, $i, j = 0$ and $M_{i,j,I,J} = 0$ for $i > 0$ or $j > 0$. This linear master equation describes the evolution of *monomer* species $M_{i,j,I,J}(t)$ in time. Such a description is different from the commonly used aggregation equation, which is a non-linear master equation that describes the evolution of the *polymer* species of different size. The most important information we like to extract from this master equation is the time dependant degree distribution $u(i, j, t)$. The latter is readily obtained by lumping together all monomer species having the same numbers of in- and out-edges:

$$
u(i, j, t) = \sum_{I,J \geq 0} M_{i,j,I,J}(t).
\tag{10}
$$

The master eq. (10) is a linear differential-difference equation that can be transformed to an analytically solvable system of partial differential equations by applying the generating function (GF) transform. We thus directly proceed by writing the expression for the degree distribution:

$$
u(i, j, t) = \sum_{I,J \geq 0} \binom{I}{i} p_A(t)^i (1 - p_A(t))^{I-i} \binom{J}{j} p_B(t)^j (1 - p_B(t))^{J-j} P(I, J).
\tag{11}
$$

where

$$
p_A(t) = \frac{\mu(t)}{\nu_{10}}, \quad \text{and} \quad p_B(t) = \frac{\mu(t)}{\nu_{01}},
\tag{12}
$$

denote the fractions of converted in- and out-edges and

$$
\mu(t) = \mu_{10}(t) = \mu_{01}(t) = \nu_{01} - \frac{\nu_{01}(\nu_{01} - \nu_{10})}{\nu_{01} - \nu_{10} e^{t(\nu_{10} - \nu_{01})}}
\tag{13}
$$

is the expected in/out degree. Here, $\nu_{mn}$ and $\mu_{mn}(t)$ refer to the mixed moments of respectively $P(I, J)$ and $u(i, j, t)$:

$$
\begin{aligned}
\nu_{mn} &= \sum_{I,J \geq 0} I^m J^n P(I, J), \\
\mu_{mn}(t) &= \sum_{i,j \geq 0} i^m j^n u(i, j, t).
\end{aligned}
\tag{14}
$$

Note, that since the expected in- and out-degrees coincide, we have $\mu(t) = \mu_{10}(t) = \mu_{01}(t)$. Mixed moments of the degree distribution $\mu_{mn}(t)$ can be obtained in the form of analytical expressions by performing the appropriate summations of eq. (11). For instance, the list of mixed moments up to order $n + m = 2$ is as follows:

$$
\begin{aligned}
\mu_{00}(t) &= 1, \\
\mu_{10}(t) &= \mu_{01}(t) = \mu(t) = p_A(t)\nu_{10} = p_B(t)\nu_{01}, \\
\mu_{20}(t) &= p_A(t)(p_A(t)\nu_{20} - p_A(t)\nu_{10} + \nu_{10}), \\
\mu_{02}(t) &= p_B(t)(p_B(t)\nu_{02} - p_B(t)\nu_{01} + \nu_{01}), \\
\mu_{11}(t) &= p_A(t)p_B(t)\nu_{11}, \\
\mu_{22}(t) &= p_A(t)p_B(t)(p_A(t)p_B(t)(\nu_{22} - \nu_{21} - \nu_{12} + \nu_{11}) \\
& \quad + p_A(t)(\nu_{21} - \nu_{11}) + p_B(t)(\nu_{12} - \nu_{11}) + \nu_{11}).
\end{aligned}
\tag{15}
$$

Below we make use of the expressions for $\mu_{i,j}(t)$ to determine the gelation conversion, whereas $u(i, j, t)$ is linked to the distribution of molecular weights, to the average molecular weight, and to the typical shortest path length.

**Generating functions of the degree and excess degree distributions.** In this section, we omit the time dependance in the degree distribution (11), and demonstrate how one can study directed networks defined by the bivariate degree distribution as combinatorial species[55,56]. The GF of a bivariate distribution is formally given by[57]:

$$U(x, y) = \sum_{i,j \geq 0} u(i, j) x^i y^j,$$

(16)

with $|x|, |y| \leq 1$, $x, y \in \mathbb{C}$ and $U(x, y)|_{x,y=1} = 1$. The excess distributions $u_{\text{in}}(i, j)$ and $u_{\text{out}}(i, j)$, are defined as the degree distributions of nodes that are reached by randomly choosing an in- or out-edge:

$$u_{\text{in}}(i, j) = \frac{1}{\mu}(i + 1)u(i + 1, j),$$

$$u_{\text{out}}(i, j) = \frac{1}{\mu}(j + 1)u(i, j + 1).$$

(17)

The GFs of the latter distributions are given by:

$$U_{\text{in}}(x, y) = \frac{1}{\mu}\frac{\partial U(x, y)}{\partial x},$$

$$U_{\text{out}}(x, y) = \frac{1}{\mu}\frac{\partial U(x, y)}{\partial y},$$

(18)

and satisfy $U_{\text{in}}(x, y)|_{x,y=1} = 1$ and $U_{\text{out}}(x, y)|_{x,y=1} = 1$. Plugging the degree distributions (11) into eqs (16–18) gives:

$$U(x, y) = \sum_{I,J \geq 0} P(I, J)((x - 1)p_{\text{A}} + 1)^I((y - 1)p_{\text{B}} + 1)^J,$$

$$U_{\text{in}}(x, y) = \frac{1}{\mu}\sum_{I,J \geq 0} P(I, J)Ip_{\text{A}}((x - 1)p_{\text{A}} + 1)^{I-1}((y - 1)p_{\text{B}} + 1)^J,$$

$$U_{\text{out}}(x, y) = \frac{1}{\mu}\sum_{I,J \geq 0} P(I, J)Jp_{\text{B}}((x - 1)p_{\text{A}} + 1)^I((y - 1)p_{\text{B}} + 1)^{J-1}.$$

(19)

Having these expressions in hand allows us to link the moments of the degree distribution $u(i, j, t)$ as defined by eq. (14) to the partial derivatives of the GFs by writing:

$$\mu_{mn} = \left[ \left( x\frac{\partial}{\partial x} \right)^m \left( y\frac{\partial}{\partial y} \right)^n U(x, y) \right]\Bigg|_{x,y=1}.$$

(20)

These relations are used below to derive analytical expressions for various global features of the polymer network.

**Molecular weight distribution.** From chemical point of view any cluster of monomers that are connected together by means of covalent bonds is considered to be a molecule. In our directed network a molecule is therefore represented by a connected component, whereas the molecular weight is simply the size of this component. The distribution of molecular weights is a popular descriptor of polymer materials. In fact, there are two ways to define such distribution: the probability $w(s)$ that a randomly chosen *monomer* belongs to a component of size $s$ is called the molecular weight distribution. Alternatively, by applying the weight $\frac{1}{s}$ we obtain the molecular size distribution,

$$n(s) = \frac{C}{s}w(s),$$

that is the probability that a randomly chosen *molecule* has size $s$. In the latter equation $C$ provides the appropriate normalisation of probability.

Here we link the molecular weight distribution $w(s)$ to the size distribution of connected components in the directed configuration model as derived in ref.[43], and briefly discuss the insights that this interpretation brings to understanding the step-growth polymerisation polymerisation process. The first values of $w(s)$ can be found by following simple considerations. For instance $w(1)$ is the probability to choose an isolated node with no neighbours, and therefore:

$$w(1) = u(0, 0).$$

Furthermore, $w(2)$ is the probability that a randomly chosen node has one edge and its only neighbour has no edges except the one that connects it with the first node:

$$w(2) = \frac{2}{\mu}u(1,0)u(0,1).$$

Continuing this list would lead to a combinatorial explosion of possibilities. A much faster alternative is to employ the GFs. The GF for $w(s)$ is formally defined as:

$$W(x) = \sum_s w(s)x^s, \ |x| \leq 1, \ x \in \mathbb{C},$$
(21)

and is obtained from the following system of functional equations:

$$
\begin{aligned}
W(x) &= xU[W_{\text{out}}(x), W_{\text{in}}(x)], \\
W_{\text{in}}(x) &= xU_{\text{in}}[W_{\text{out}}(x), W_{\text{in}}(x)], \\
W_{\text{out}}(x) &= xU_{\text{out}}[W_{\text{out}}(x), W_{\text{in}}(x)],
\end{aligned}
$$
(22)

where the GFs $U(x,y)$, $U_{\text{in}}(x,y)$ and $U_{\text{out}}(x,y)$ are defined by eqs (16) and (18). The functions $W_{\text{in}}(x) = \sum_{s>0} w_{\text{in}}(s)x^s$ and $W_{\text{out}}(x) = \sum_{s>0} w_{\text{out}}(s)x^s$ denote the GFs of the excess component size distributions $w_{\text{in}}(s)$ and $w_{\text{out}}(s)$.

The formal solution to eq. (22) is given by the following relation[43]:

$$
w(s) = \begin{cases}
\sum_{m=0}^{s-1} a(m, s-m-1), & s > 1, \\
u(0,0), & s = 1,
\end{cases}
$$
(23)

where

$$a(m, n) = u(i,j) \ * \ u_{\text{out}}(i,j)^{*m-1} \ * \ u_{\text{in}}(i,j)^{*n-1} \ * \ d(i,j)|_{\substack{i=m \\ j=n}}, \ m, \ n \geq 0,$$
(24)

and

$$d(i,j) = [u_{\text{out}}(i,j) - iu_{\text{out}}(i,j)] \ * \ [u_{\text{in}}(i,j) - ju_{\text{in}}(i,j)] - ju_{\text{out}}(i,j) \ * \ iu_{\text{in}}(i,j).$$
(25)

Here, $f(i,j)^{*n}$ denotes the convolution power $f(i,j)^{*n} = f(i,j)^{*n-1} \ * \ f(i,j)$, $f(i,j)^{*0} := 1$, and the bivariate convolution is defined as

$$f(i,j) \ * \ g(i,j) = \sum_{i_1+i_2=i, j_1+j_2=j} f(i_1, j_1)g(i_2, j_2).$$
(26)

In practice, numerical values of the convolution can be conveniently obtained by Fast Fourier Transform FFT. The asymptotical analysis of $w(s)$ yields two distinct asymptotes:

$$w_\infty(s) \propto s^{-\frac{3}{2}} e^{-C_1 s}, \ \text{if } \nu_{20} - \nu_{10} > 0 \ \text{or} \ \nu_{02} - \nu_{10} > 0,$$
(27)

and

$$w_\infty(s) \propto s^{-\frac{1}{2}} e^{-C'_1 s}, \ \text{if } \nu_{20} - \nu_{10} = 0 \ \text{or} \ \nu_{02} - \nu_{10} = 0.$$
(28)

The exact expression for the coefficients $C_1$ and $C_{1'}$ are given in ref.[43] and $\nu_{nm}$ are as defined in eq. (14).

**Systems with a single monomer type.** Although the numerical values of eq. (23) are accessible at the cost of $\mathcal{O}(s^2 \log s)$ operations, explicit analytical relations can be obtained in some special cases. Consider the case when there is only one monomer species bearing $I$ groups of type A and $J$ groups of type B, that is the $A_I B_J$ monomer. We will now derive an explicit analytical expression for $w(s)$. Note that in this case, the distribution of initial functionalities is trivial, $P(I,J) = 1$, and therefore the expression for the degree distribution given in eq. (11) simplifies to a bivariate binomial distribution:

$$u(i,j,t) = \binom{I}{i}p_A(t)^i(1-p_A(t))^{I-i}\binom{J}{j}p_B(t)^j(1-p_B(t))^{J-j},$$
(29)

where

$$p_A(t) = J\frac{e^{t(I-J)} - 1}{Ie^{t(I-J)} - J} \ \text{and} \ p_B(t) = I\frac{e^{t(I-J)} - 1}{Ie^{t(I-J)} - J}.$$
(30)

The first values of $w(s)$ are readily obtained by writing:

$$w(1) = u(0,0) = (1 - p_A)^I (1 - p_B)^J,$$

$$w(2) = \frac{2}{\mu} u(1,0) u(0,1) = 2J p_B (1 - p_A)^{2I-1} (1 - p_B)^{2J-1}$$

$$= 2I p_A (1 - p_A)^{2I-1} (1 - p_B)^{2J-1}.$$

Since $u(i,j)$ has a binomial form, one can analytically solve the convolution powers appearing in eq. (24) to obtain:

$$w(s) = p_B^{s-2} (1 - p_A)^{(I-1)s+1} (1 - p_B)^{(J-1)s+1}$$

$$\times \left[ p_B \binom{Js}{s-1} A - p_B J \binom{Js-1}{s-2} B - p_A (I-1) \binom{Js-2}{s-2} C \right], \tag{31}$$

where factors $A$, $B$, and $C$ are defined via the Hypergeometric function:

$$A = {}_2F_1 \left( 1 - s, (I-1)s + 1; -Js; \frac{p_A}{p_B} \right),$$

$$B = {}_2F_1 \left( 2 - s, (I-1)s + 1; 1 - Js; \frac{p_A}{p_B} \right),$$

$$C = {}_2F_1 \left( 2 - s, (I-1)s + 2; 2 - Js; \frac{p_A}{p_B} \right).$$

As a quick check-up, we consider the $AB_2$-system by setting $I = 1, J = 2$ and $p = \frac{p_A}{2} = p_B \leq 0.5$. This substitution reduces eq. (31) to:

$$w(s) = s(1 - 2p) p^{s-1} (1 - p)^{s+1} \frac{(2s)!}{s!(s+1)!}. \tag{32}$$

This result is identical, up to the normalisation factor $s$, to the size distribution derived in ref.[11]. Note we normalise with $s$ to guarantee that $w(s)$ is a proper probability mass function, that is $\sum_{s \geq 1} w(s) = 1$.

**Weight-average molecular weight.** The weight-average molecular weight is a widely-used quantity in polymer chemistry. It is defined by the follwing ratio:

$$\langle s \rangle := \frac{\sum_{s>0} s w(s)}{\sum_{s>0} w(s)} = \frac{W'(1)}{W(1)}. \tag{33}$$

Note that before the gel point, $w(s)$, $w_{in}(s)$ and $w_{out}(s)$ are appropriately normalised,

$$\sum_{s>0} w(s) = W(1) = W_{in}(1) = W_{out}(1) = 1.$$

By plugging eq. (22) into definition (33) we obtain:

$$\langle s \rangle_{t<t_{gel}} = \frac{W'(1)}{W(1)} = W'(1) = \frac{\mu^2(-2\mu_{11} + \mu_{20} + \mu_{02})}{\mu_{11}^2 - 2\mu\mu_{11} - \mu_{02}\mu_{20} + \mu(\mu_{20} + \mu_{02})} + 1. \tag{34}$$

After gel transition, the latter expression becomes more complex and reads:

$$\langle s \rangle_{t>t_{gel}} = \frac{W'(1)}{U(r_{out}, r_{in})} = \frac{\mu^2}{U(r_{out}, r_{in})} \cdot \frac{2r_{in}r_{out}(\mu - U_{11}(r_{out}, r_{in})) + r_{in}^2 U_{02}(r_{out}, r_{in}) + r_{out}^2 U_{20}(r_{out}, r_{in})}{\mu^2 - 2\mu U_{11}(r_{out}, r_{in}) - U_{02}(r_{out}, r_{in})U_{20}(r_{out}, r_{in}) + U_{11}^2(r_{out}, r_{in})} + 1, \tag{35}$$

where $U_{lm}(x,y) = \left(\frac{\partial}{\partial x}\right)^l \left(\frac{\partial}{\partial y}\right)^m U(x,y)$ denote the partial derivatives, and $(r_{in}, r_{out})$ is the solution of the following system of equations:

$$r_{in} = U_{in}(r_{out}, r_{in}),$$

$$r_{out} = U_{out}(r_{out}, r_{in}). \tag{36}$$

**Phase transition and gel fraction.** During the evolution of the network the functional groups are converted into edges and at some critical point the system accumulates so many edges that it percolates. This critical moment can be identified by a few alternative methods. For instance, one may study the asymptotical behaviour of the size distribution of connected components as given by (27). This asymptote becomes scale-free at the critical point. The other alternative is to directly detect the percolation phase transition by looking at the degree distribution itself. In this case, the changes that occur in the degree distribution at the critical point are more subtle, yet they can be detected by a specially designed criticality criterion. This criterion was introduced by Molloy and

Reed for undirected networks[47], and was later generalised to the case of directed networks in ref.[29]. In this section we briefly discuss the implications of the latter theory on our dynamic polymer network.

If the only available information about a system is its degree distribution, we can detect whether the system is in the gel regime by the following criterion:

$$\mu_{11}^2 - 2\mu\mu_{11} - \mu_{02}\mu_{20} + \mu(\mu_{20} + \mu_{02}) \leq 0. \tag{37}$$

The conversion of A-groups at the critical point is given by:

$$p_{\text{A,critical}} = \frac{\nu_{01}}{\nu_{11} + \sqrt{(\nu_{02} - \nu_{01})(\nu_{20} - \nu_{10})}}. \tag{38}$$

Thus, if $p_A(t) > p_{\text{A,critical}}$ the system contains gel. Some system, however, never produce gel. This happens because the initial configuration of the system does not have a sufficient amount of high functional monomers, or too many monomers of functionality one that terminate the growth of the network. In either case this statement can be quantified by looking at the moments of the functionality distribution $P(I, J)$: the phase transition occurs in finite time if at least one of the following conditions is true:

$$(\nu_{02} - \nu_{01})(\nu_{20} - \nu_{10}) - (\nu_{11} - \nu_{10})^2 > 0, \text{ and } \nu_{01} \leq \nu_{10}.$$
$$\text{or}$$
$$(\nu_{02} - \nu_{01})(\nu_{20} - \nu_{10}) - (\nu_{11} - \nu_{01})^2 > 0, \text{ and } \nu_{01} \geq \nu_{10}, \tag{39}$$

The gel fraction is defined as the probability that a randomly selected node belongs to the gel molecule. The GF of the component size distribution $W(x)$ only describes the components of finite size, and the gel fraction is found as the mass deficit that departs from zero at the phase transition. The amount of this 'lost' mass, that is the probability that a randomly chosen monomer belongs to the gel, is given by

$$g_f = 1 - r, \tag{40}$$

where $r = W(1)$. This means that in order to recover $g_f$, one needs to solve the equation for $W(x)$ only at a single point $x = 1$. By substituting $x = 1$ into (22) one obtains:

$$\begin{aligned} r &= U(r_{\text{out}}, r_{\text{in}}), \\ r_{\text{in}} &= U_{\text{in}}(r_{\text{out}}, r_{\text{in}}), \\ r_{\text{out}} &= U_{\text{out}}(r_{\text{out}}, r_{\text{in}}), \end{aligned} \tag{41}$$

where $U(x, y)$, $U_{\text{in}}(x, y)$ and $U_{\text{out}}(x, y)$ are given by eq. (19).

It is important to note that the directionality of the network, as it is introduced in the present paper, is decisive only in the case of an asymmetric degree distribution $u(i, j) \neq u(j, i)$. In case of a symmetric distribution $u(i, j) = u(j, i)$, the directed network model will give the same results as the undirected model supplied with degree distribution $u(k) = \sum_{k=i+j} u(i, j)$. This implies that an AB-system (undergoing AB-reactions) with symmetric initial input $(P(I, J) = P(I, J))$ leads to the same network as an AA-system (undergoing AA-reactions) with $P(K) = \sum_{K=I+J} P(I, J)$. For example, the AB-system $A_3:B_3 = 1:1$ forms a topologically identical polymer network (when disregarding the directionality of the edges) as the AA-system $A_3$. With this in mind, eq. (38) simplifies for the case of symmetric functionality distribution $P(I, J) = P(I, J)$ as follows:

1. Symmetry leading to

$$p_{\text{critical}} = \frac{\nu_{10}}{\nu_{11} + \nu_{20} - \nu_{10}}, \tag{42}$$

with $\nu_{01} = \nu_{10}$ and $\nu_{02} = \nu_{20}$.

2. Mapping the symmetric AB-system to an AA-system (e.g. $A_3:B_3 = 1:1 \equiv A_3$, $A_2B:AB_2 = 1:1 \equiv A_3$) simplifies to

$$p_{\text{critical}} = \frac{\nu_1}{\nu_2 - \nu_1}, \tag{43}$$

using $\nu_{10} = \frac{1}{2}\nu_1$ and $\nu_{11} + \nu_{20} = \frac{1}{2}\nu_2$. The deduced equations give the critical conversion in undirected networks[22].

3. For systems of one initial monomer type we note $\nu_2 = \nu_1^2$, giving

$$p_{\text{critical}} = \frac{1}{\nu_1 - 1} = \frac{1}{I + J - 1}. \tag{44}$$

This corresponds to the Flory-Stockmayer equation, where $z = \nu_1$ is the expected number of first neighbours[9].

**Derivation of the gyration radius.** In polymer physics, the radius of gyration is used to describe the dimensions of a branched polymer and can be experimentally observed by light scattering experiments. Consider

a branched polymer with $s$ monomers having coordinates $r_i \in \mathbb{R}^3$, $i = 1, \ldots, s$. The radius of gyration $R(s)_g^2$ of this topology is conventionally defined as:

$$R_g^2(s) = \frac{1}{s^2} \sum_{k=1}^{s} \sum_{l=k}^{s} (\overrightarrow{r_k} - \overrightarrow{r_l})^2. \tag{45}$$

This quantity can be estimated using the Kramer's theorem[58] that states:

$$\frac{R_g^2(s)}{b^2} = \frac{1}{s^2} \sum_{j=1}^{s-1} s_L(j) s_R(j), \tag{46}$$

where $b$ is the Kuhn's length, which is related to the size of a monomer unit. This sum runs over all possible cuts of the branched structure into two fragments: the left fragment of size $s_L$ and the right fragment of size $s_R$. There are $s-1$ of such partitions. In a statistical ensemble, the size distributions of $s_L$ and $s_R$ are given by respectively $w_{in}(s)$ and $w_{out}(s)$, which are defined by their GFs in eq. (22). Hillegers & Sloot[38] formulated the ensemble average for eq. (46) with respect to $w_{in}(s)$ and $w_{out}(s)$:

$$\frac{\langle R_g^2(s) \rangle}{b^2} = \frac{1}{s^2} \frac{\sum_{s_A + s_B = s} s_A w_{in}(s_A) s_B w_{out}(s_B)}{\frac{1}{s-1} \sum_{s_A + s_B = s} w_{in}(s_A) w_{out}(s_B)}, \tag{47}$$

which we further process using a discrete Fourier transform $\mathcal{F}^{-1} G(k)|_s$:

$$\frac{\langle R_g^2(s) \rangle}{b^2} = \frac{s-1}{s^2} \frac{(s w_{in}(s)) * (s w_{out}(s))}{w_{in}(s) * w_{out}(s)} = \frac{s-1}{s^2} \frac{\mathcal{F}^{-1}(W'_{in}(x_k) W'_{out}(x_k))|_{s-2}}{\mathcal{F}^{-1}(W_{in}(x_k) W_{out}(x_k))|_s}, \tag{48}$$

with $x_k = e^{-2\pi i \frac{k}{N+1}}$, $k = 0, \ldots, N$. Therefore, if $W_{in}(x)$ and $W_{out}(x)$ are already available from the computation of the component size distribution $w(s)$, the ensemble-average radius of gyration is obtained by applying the FFT algorithm at the cost of $\mathcal{O}(s \log s)$ multiplicative operations. For small $s$, it is possible to compute the convolution directly, whereas eq. (48) is the most advantageous for $s \gg 1$, where the direct evaluation of the convolution becomes unfeasible.

Another related quantity to the gyration radius is the contraction factor $g(s)$, which is given by the ratio between the mean square gyration radius for a branched polymer $R_g^2(s)$ and the square gyration radius of a reference linear polymer with the same length $R_g^2(s)|_{linear}$:

$$g(s) = \frac{R_g^2(s)}{R_g^2(s)|_{linear}}, \tag{49}$$

where $R_g^2(s)|_{linear}$ is typically estimated from the Gaussian coil model:

$$\left. \frac{R_g^2(s)}{b^2} \right|_{linear} = \frac{s^2 - 1}{6s}. \tag{50}$$

**Scaling of the node neighbourhood size.**    In this section we apply the Joyal's theory of combinatorial species to investigate the number of nodes that are contained within a given topological distance from a randomly chosen node. The expected number of first-degree neighbours is defined as the sum of the neighbours reached by the in-edges and the neighbours reached by the out-edges. By using GF this number can be extracted from the degree distribution:

$$z_1 = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) U(x, y)|_{x,y=1} = 2\mu, \tag{51}$$

and moreover, the expected number of the $m^{th}$-degree neighbours is given by a composition of $U(x, y)$ and $(m-1)$-fold composition of the excess GF:

$$z_m = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) U^{[m]}(x, y)|_{x,y=1}. \tag{52}$$

Here $U^{[m]}$ generates the probability for the number of $m^{th}$ neighbours:

$$U^{[m]} := \begin{cases} U(x, y), & \text{for } m = 1, \\ U^{[m-1]}(U_{out}(x, y), U_{in}(x, y)), & \text{for } m > 1. \end{cases} \tag{53}$$

Therefore, for the first-degree neighbours we have $z_1 = 2\mu$. For second-degree neighbours we have:

$$z_2 = \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) U(U_{\text{out}}(x, y),\ U_{\text{in}}(x, y))|_{x,y=1}$$

$$= \mu\left(\frac{1}{\mu}\frac{\partial^2}{\partial^2 x} + 2\frac{1}{\mu}\frac{\partial^2}{\partial x \partial y} + \frac{1}{\mu}\frac{\partial^2}{\partial^2 y}\right) U(x, y)|_{x,y=1}$$

$$= \mu \mathbf{I}^T\left(\frac{1}{\mu}\mathbf{L}U(x, y)|_{x,y=1}\right)\mathbf{1}, \tag{54}$$

where $\mathbf{L} = \begin{pmatrix} \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial^2 y} \\ \frac{\partial^2}{\partial^2 x} & \frac{\partial^2}{\partial x \partial y} \end{pmatrix}$ and $\mathbf{I}^T = (11)$. The expected number of the third-degree neighbours is given by:

$$z_3 = \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) U(U_{\text{out}}(U_{\text{out}}(x, y),\ U_{\text{in}}(x, y)),\ U_{\text{in}}(U_{\text{out}}(x, y),\ U_{\text{in}}(x, y)))|_{x,y=1}$$

$$= \mu\Bigg(\frac{1}{\mu^2}\frac{\partial^2}{\partial x \partial y}\left(\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial^2 y}\right) + \frac{1}{\mu^2}\frac{\partial^2}{\partial^2 y}\left(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial x \partial y}\right)$$

$$+ \frac{1}{\mu^2}\frac{\partial^2}{\partial^2 x}\left(\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial^2 y}\right) + \frac{1}{\mu^2}\frac{\partial^2}{\partial x \partial y}\left(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial x \partial y}\right)\Bigg) U(x, y)|_{x,y=1}$$

$$= \mu \mathbf{I}^T\left(\frac{1}{\mu}\mathbf{L}U(x, y)|_{x,y=1}\right)^2 \mathbf{1}. \tag{55}$$

By using induction we arrive at the following expression for the expected number of the $m^{\text{th}}$-degree neighbours in terms of the degree distribution moments:

$$z_m = \mu \mathbf{I}^T \mathbf{A}^{m-1}\mathbf{1}, \tag{56}$$

where

$$\mathbf{A} = \frac{1}{\mu}\mathbf{L}U(x, y)|_{x,y=1} = \frac{1}{\mu}\begin{pmatrix} \mu_{11} & \mu_{02} - \mu \\ \mu_{20} - \mu & \mu_{11} \end{pmatrix}. \tag{57}$$

Now, the total number of nodes $N$ contained within a topological ball of radius $l$ is given as a sum:

$$N = 1 + \sum_{m=1}^{l} z_m = 1 + \mu\sum_{m=1}^{l} \mathbf{I}^T \mathbf{A}^{m-1}\mathbf{1} = 1 + \mu\mathbf{I}^T\left(\sum_{m=1}^{l} \mathbf{A}^{m-1}\right)\mathbf{1}. \tag{58}$$

Using the equation for the sum of the geometric series, $\sum_{m=0}^{l} \mathbf{A}^m = (\mathbf{A}^{l+1} - \mathbf{I})(\mathbf{A} - \mathbf{I})^{-1}$, with $I$ denoting the identity matrix, we obtain:

$$N = 1 + \mu\mathbf{I}^T(\mathbf{A}^l - \mathbf{I})(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1}. \tag{59}$$

The latter equality transforms to:

$$\mathbf{I}^T\mathbf{A}^l(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1} = \frac{N - 1}{\mu} + \mathbf{I}^T(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1}. \tag{60}$$

We will now perform an asymptotic analysis of the latter equation by assuming that $l \gg 1$, in which case, the asymptotic behaviour of $\mathbf{A}^l$ is driven by the leading eigenvalue of $\mathbf{A}$. Using the eigendecomposition $\mathbf{A}^l = \mathbf{P}\mathbf{D}^l\mathbf{P}^{-1}$ gives $\mathbf{I}^T\mathbf{P}\mathbf{D}^l\mathbf{P}^{-1}(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1} = \frac{N-1}{\mu} + \mathbf{I}^T(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1}$, and defining $\mathbf{a}^T = \mathbf{1}^T\mathbf{P}$, $\mathbf{b} = \mathbf{P}^{-1}(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1}$ and $\mathcal{C} = \mathbf{I}^T(\mathbf{A} - \mathbf{I})^{-1}\mathbf{1}$ leads to

$$\mathbf{a}^T\mathbf{D}^l\mathbf{b} = \frac{N - 1}{\mu} + \mathcal{C},$$

or equivalently,

$$a_1 b_1 \lambda_1^l + a_2 b_2 \lambda_2^l = \frac{N - 1}{\mu} + \mathcal{C}, \tag{61}$$

where the eigenvalues of the matrix $\mathbf{A}$, $\lambda_{1,2} = \frac{1}{\mu}\left(\mu_{11} \pm \sqrt{(\mu_{20} - \mu)(\mu_{02} - \mu)}\right)$, are defined by the characteristic polynomial:

$$\left| \left( \frac{1}{\mu} \begin{matrix} \mu_{11} & \mu_{02} - \mu \\ \mu_{20} - \mu & \mu_{11} \end{matrix} \right) - \lambda \mathbf{I} \right| = 0. \tag{62}$$

Note that the gel criterion given by eq. (37) can also be rewritten as a determinant:

$$\det \mathbf{A}' = \left| \frac{1}{\mu} \begin{pmatrix} \mu_{11} & \mu_{02} - \mu \\ \mu_{20} - \mu & \mu_{11} \end{pmatrix} - \mathbf{I} \right| \le 0. \tag{63}$$

Thus, at the gel point, the matrix $\mathbf{A}'$ has at least one eigenvalue equal to zero. The relation of the eigenvalues of matrix $\mathbf{A}'$, $\lambda'_1$ and $\lambda'_2$, to the eigenvalues of $\mathbf{A}$, $\lambda_1$ and $\lambda_2$, is as follows: $\lambda'_1 = \lambda_1 - 1$ and $\lambda'_2 = \lambda_2 - 1$. Furthermore, above the gel transition, $\det \mathbf{A}' = \lambda'_1 \lambda'_2 < 0$. This is the case, only if one eigenvalue, $\lambda'_1 > 0$, is positive and the other one, $\lambda'_2 < 0$, is negative, and therefore, the eigenvalues of $\mathbf{A}$ satisfy $\lambda_1 > 1$, $\lambda_2 < 1$ and $|\lambda_1| > |\lambda_2|$. The implications for eq. (61) are as follows: for large $l$, $\lambda_1^l \gg \lambda_2^l$, and consequently, the total number of nodes $N$ contained within a topological ball of radius $l$ features the exponential growth after the gel transition:

$$N \propto \lambda_1^l = e^{l/l_0}, \quad l_0 = (\log \lambda_1)^{-1}, \quad l \gg 1. \tag{64}$$

## Data Availability

No datasets were generated or analysed during the current study.

## References

1. Newman, M. E. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).
2. Strogatz, S. H. Exploring complex networks. *nature* **410**, 268 (2001).
3. Moore, C. & Newman, M. E. Epidemics and percolation in small-world networks. *Phys. Rev. E* **61**, 5678 (2000).
4. Gomez, S. *et al*. Diffusion dynamics on multiplex networks. *Phys. review letters* **110**, 028701 (2013).
5. De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **12**, 901 (2016).
6. Callaway, D. S., Newman, M. E., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Phys. review letters* **85**, 5468 (2000).
7. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186 (2009).
8. Newman, M. E. The structure of scientific collaboration networks. *Proc. national academy sciences* **98**, 404–409 (2001).
9. Flory, P. J. Molecular size distribution in three dimensional polymers. I. Gelation1. *J. Am. Chem. Soc.* **63**, 3083–3090 (1941).
10. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–60 (1960).
11. Ziff, R. M. Kinetics of polymerization. *J. Stat. Phys.* **23**, 241–263 (1980).
12. Ziff, R. & Stell, G. Kinetics of polymer gelation. *The J. Chem. Phys.* **73**, 3492–3499 (1980).
13. Bunde, A., Havlin, S. & Porto, M. Are branched polymers in the universality class of percolation? *Phys. review letters* **74**, 2714 (1995).
14. Havlin, S., Djordjevic, Z. V., Majid, I., Stanley, H. & Weiss, G. Relation between dynamic transport properties and static topological structure for the lattice-animal model of branched polymers. *Phys. review letters* **53**, 178 (1984).
15. Majid, I., Jan, N., Coniglio, A. & Stanley, H. E. Kinetic growth walk: A new model for linear polymers. *Phys. review letters* **52**, 1257 (1984).
16. Stanley, H. E. Application of fractal concepts to polymer statistics and to anomalous transport in randomly porous media. *J. Stat. Phys.* **36**, 843–860 (1984).
17. Krapivsky, P. L. & Redner, S. Organization of growing random networks. *Phys. Rev. E* **63**, 066123, https://doi.org/10.1103/PhysRevE.63.066123 (2001).
18. Buldyrev, S. *et al*. Anomalous interface roughening in porous media: Experiment and model. *Phys. Rev. A* **45**, R8313 (1992).
19. Bunde, A. & Havlin, S. *Fractals and disordered systems*. (Springer Science & Business Media, 2012).
20. Stanley, H. E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
21. Papadopoulos, L., Porter, M. A., Daniels, K. E. & Bassett, D. S. Network analysis of particles and grains. *J. Complex Networks* cny005, https://doi.org/10.1093/comnet/cny005 (2018).
22. Kryven, I. Analytic results on the polymerisation random graph model. *J. Math. Chem.* **56**, 140–157 (2018).
23. Orlova, Y., Kryven, I. & Iedema, P. D. Automated reaction generation for polymer networks. *Comput. & Chem. Eng.* **112**, 37–47 (2018).
24. Schamboeck, V., Kryven, I. & Iedema, P. D. Acrylate network formation by free-radical polymerization modeled using random graphs. *Macromol. Theory Simulations* **26** (2017).
25. Ye, D. *et al*. Scaling of average receiving time on weighted polymer networks with some topological properties. *Sci. Reports* **7**, 2128 (2017).
26. Yang, Y., Qiu, F., Zhang, H. & Yang, Y. The rouse dynamic properties of dendritic chains: A graph theoretical method. *Macromolecules* **50**, 4007–4021 (2017).
27. Kryven, I., Duivenvoorden, J., Hermans, J. & Iedema, P. D. Random graph approach to multifunctional molecular networks. *Macromol. Theory Simulations* **25**, 449–465 (2016).
28. Odian, G. *Principles of polymerization* (John Wiley & Sons, 2004).
29. Kryven, I. Emergence of the giant weak component in directed random graphs with arbitrary degree distributions. *Phys. Rev. E* **94**, 012315 (2016).
30. Stockmayer, W. H. Theory of molecular size distribution and gel formation in branched polymers ii. general cross linking. *The J. Chem. Phys.* **12**, 125–131 (1944).
31. Durand, D. & Bruneau, C. M. General expressions of average molecular weights in condensation polymerisation of polyfunctional monomers. *Polym. Int.* **11**, 194–198 (1979).
32. Matsoukas, T. Statistical thermodynamics of irreversible aggregation: the sol-gel transition. *Sci. reports* **5**, 8855 (2015).
33. Lushnikov, A. An exactly solvable model of polymerization. *Chem. Phys.* **493**, 133–139 (2017).
34. Kryven, I. & Iedema, P. D. A novel approach to population balance modeling of reactive polymer modification leading to branching. *Macromol. Theory Simulations* **22**, 89–106 (2013).

35. Kryven, I. & Iedema, P. D. Transition into the gel regime for crosslinking radical polymerisation in a continuously stirred tank reactor. *Chem. Eng. Sci.* **126**, 296–308 (2015).
36. Müller, A. H., Yan, D. & Wulkow, M. Molecular parameters of hyperbranched polymers made by self-condensing vinyl polymerization. 1. Molecular weight distribution. *Macromolecules* **30**, 7015–7023 (1997).
37. Tobita, H. Universality in branching frequencies and molecular dimensions during hyperbranched polymer formation: 2. Step polymerization of ab2 type monomer with different reactivity for the second b group. *Macromol. Theory Simulations* **25**, 123–133 (2016).
38. Hillegers, L. T. & Slot, J. J. Step-growth polymerizing systems of general type "afibgi": Calculating the radius of gyration and the g-curve using generating functions and recurrences. *Macromol. Theory Simulations* **26** (2017).
39. Torres-Knoop, A., Kryven, I., Schamboeck, V. & Iedema, P. Modeling the free-radical polymerization of hexanediol diacrylate (hdda): a molecular dynamics and graph theory approach. *Soft matter* (2018).
40. Izumi, A., Shudo, Y., Hagita, K. & Shibayama, M. Molecular dynamics simulations of cross-linked phenolic resins using a united-atom model. *Macromol. Theory Simulations* 1700103 (2018).
41. Newman, M. E., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. review E* **64**, 026118 (2001).
42. Kryven, I. General expression for the component size distribution in infinite configuration networks. *Phys. Rev. E* **95**, 052303 (2017).
43. Kryven, I. Finite connected components in infinite directed and multiplex networks with arbitrary degree distributions. *Phys. Rev. E* **96**, 052304 (2017).
44. Metri, V. *et al.* Physical networks from multifunctional telechelic star polymers: A rheological study by experiments and simulations. *Macromolecules* (2018).
45. Yan, T., Schröter, K., Herbst, F., Binder, W. H. & Thurn-Albrecht, T. Unveiling the molecular mechanism of self-healing in a telechelic, supramolecular polymer network. *Sci. reports* **6**, 32356 (2016).
46. de Greef, T. F. & Meijer, E. Materials science: supramolecular polymers. *Nature* **453**, 171 (2008).
47. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random structures & algorithms* **6**, 161–180 (1995).
48. Ben-Naim, E. & Krapivsky, P. Kinetic theory of random graphs: From paths to cycles. *Phys. Rev. E* **71**, 026129 (2005).
49. Ben-Naim, E. & Krapivsky, P. Dynamics of random graphs with bounded degrees. *J. Stat. Mech. Theory Exp.* **2011**, P11008 (2011).
50. Chen, H. & Kong, J. Terminal index: a new way for precise description of topologic structure of highly branched polymers derived from a2 + b3 stepwise polymerization. *The J. Phys. Chem. B* **118**, 3441–3450 (2014).
51. Stauffer, D. & Aharony, A. *Introduction to percolation theory: revised second edition* (CRC press, 2014).
52. Burchard, W. Static and dynamic light scattering from branched polymers and biopolymers. In *Light scattering from polymers*, 1–124 (Springer, 1983).
53. Tacx, J. C. & Iedema, P. D. Simulating light scattering behavior of branched molecules. *Macromol. Theory Simulations* **26** (2017).
54. Dobrynin, A. A., Entringer, R. & Gutman, I. Wiener index of trees: theory and applications. *Acta Appl. Math* **66**, 211–249 (2001).
55. Joyal, A. Une théorie combinatoire des séries formelles. *Adv. mathematics* **42**, 1–82 (1981).
56. Bergeron, F. *et al. Combinatorial species and tree-like structures*, vol. 67 (Cambridge University Press, 1998).
57. Wilf, H. S. *Generating functionology* (AK Peters/CRC Press, 2005).
58. Rubinstein, M. & Colby, R. H. *Polymer physics*, vol. 23 (Oxford university press New York, 2003).
59. McKee, M. G., Unal, S., Wilkes, G. L. & Long, T. E. Branched polyesters: recent advances in synthesis and performance. *Prog. Polym. Sci.* **30**, 507–539 (2005).
60. Mahapatra, S. S., Yadav, S. K., Yoo, H. J., Cho, J. W. & Park, J.-S. Highly branched polyurethane: Synthesis, characterization and effects of branching on dispersion of carbon nanotubes. *Compos. Part B: Eng.* **45**, 165–171 (2013).
61. Tuerp, D. & Bruchmann, B. Dendritic polyurea polymers. *Macromol. rapid communications* **36**, 138–150 (2015).
62. Chao, D. *et al.* Multifunctional hyperbranched polyamide: synthesis and properties. *Polymer* **54**, 3223–3229 (2013).

## Acknowledgements

## Author Contributions

I.K. designed research and developed the theoretical formalism, V.S. derived analytical equations, performed numerical simulations, and composed figures, V.S. and P.D.I. performed the literature study, wrote Introduction and Results, V.S. and I.K. wrote Methods. All authors discussed the results and contributed to writing the final manuscript.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.