PERCOLATION

A Mathematical Way of Knowing Epidemics

`Following Newman (2002): Spread of epidemic disease on networks`

Hiroki Kojima — Chiba Institute of Technology

Ways of Knowing: Spread of Infectious Disease in Urban Environments

PHYSICAL REVIEW E **66**, 016128 (2002)

Spread of epidemic disease on networks

M. E. J. Newman
Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109-1120
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
(Received 4 December 2001; published 26 July 2002)

The study of social networks, and in particular the spread of disease on networks, has attracted considerable recent attention in the physics community. In this paper, we show that a large class of standard epidemiological models, the so-called susceptible/infective/removed (SIR) models can be solved exactly on a wide variety of networks. In addition to the standard but unrealistic case of fixed infectiveness time and fixed and uncorrelated probability of transmission between all pairs of individuals, we solve cases in which times and probabilities are nonuniform and correlated. We also consider one simple case of an epidemic in a structured population, that of a sexually transmitted disease in a population divided into men and women. We confirm the correctness of our exact solutions with numerical simulations of SIR epidemics on networks.

DOI: 10.1103/PhysRevE.66.016128 PACS number(s): 89.75.Hc, 87.23.Ge, 05.70.Fh, 64.60.Ak

I. INTRODUCTION

Many diseases spread through human populations by contact between infective individuals (those carrying the disease) and susceptible individuals (those who do not have the disease yet, but can catch it). The pattern of these disease-causing contacts forms a network. In this paper we investigate the effect of network topology on the rate and pattern of disease spread.

Most mathematical studies of disease propagation make the assumption that populations are "fully mixed," meaning

within the statistical physics community, has addressed the topological properties of networks of various kinds, from both theoretical and empirical points of view, and studied the effects of topology on processes taking place on those networks [4,5]. Social networks [6–9], technological networks [10–13], and biological networks [14–18] have all been examined and modeled in some detail. Building on insights gained from this work, a number of authors have pursued a mathematical theory of the spread of disease on networks [19–24]. This is also the topic of the present paper, in which we show that a large class of standard epidemiological mod-

SIR model

[+/−] Speed [1/2/3] Scenarios

The textbook lens: SIR

Before today's two lenses — the textbook baseline. Open almost any epidemiology textbook:

$$\frac{dS}{dt} = -\frac{\beta S I}{N}$$
$$\frac{dI}{dt} = \frac{\beta S I}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$
$$R_0 = \frac{\beta}{\gamma}$$

S, I, R	**S**usceptible · **I**nfectious · **R**ecovered — number of people in each compartment at time t
N	total population (= S + I + R, held constant — no births or deaths during the model run)
β	transmission rate — expected new infections per infectious–susceptible contact, per unit time
γ	recovery rate — equals *1 / (average infectious period)*; large γ means people recover quickly
R_0	basic reproduction number — β / γ, the average new infections one sick person causes in a fully-susceptible population

M. E. J. NEWMAN

then they become infected. In the limit of large N this model is governed by the coupled nonlinear differential equations [1]:

$$\frac{ds}{dt} = -\beta i s, \quad \frac{di}{dt} = \beta i s - \gamma i, \quad \frac{dr}{dt} = \gamma i, \tag{1}$$

where $s(t)$, $i(t)$, and $r(t)$ are the fractions of the population in each of the three states, and the last equation is redundant, since $s + i + r = 1$ necessarily at all times. This model is appropriate for a rapidly spreading disease that confers immunity on its survivors, such as influenza. In this paper we will consider only diseases of this type. Diseases that are endemic because they propagate on time scales comparable to or slower than the rate of turnover of the population, or because they confer only temporary immunity, are not well represented by this model; other models have been developed for these cases [3].

The model described above assumes that the population is fully mixed, meaning that the individuals with whom a sus-

SIR on a Random Network

Random network

N nodes. Degree distribution p_k. Configuration model → locally tree-like (few short loops).

SIR dynamics

S → I → R. Infective node i transmits to neighbor j with rate r_{ij} per day, for τ_i days.

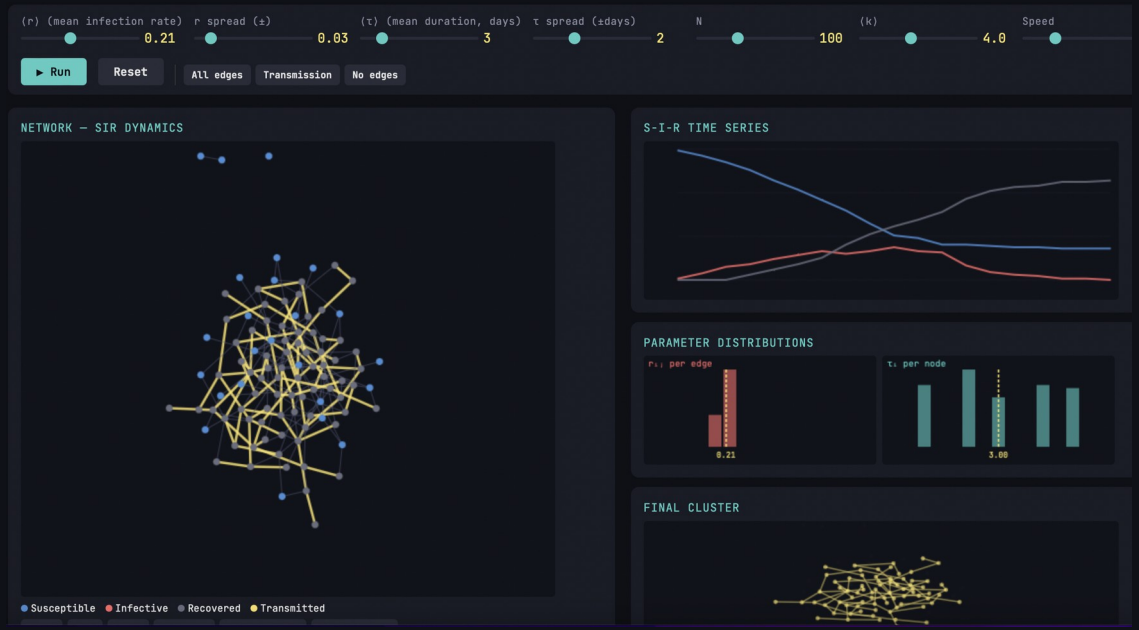
Per-edge, per-node randomness

$r_{ij} \sim P(r)$, $\tau_i \sim P(\tau)$ — both i.i.d. random.

Edge transmission probability: $T_{ij} = 1 - (1-r_{ij})^{\tau_i}$

LIVE DEMO

Watch the Dynamics

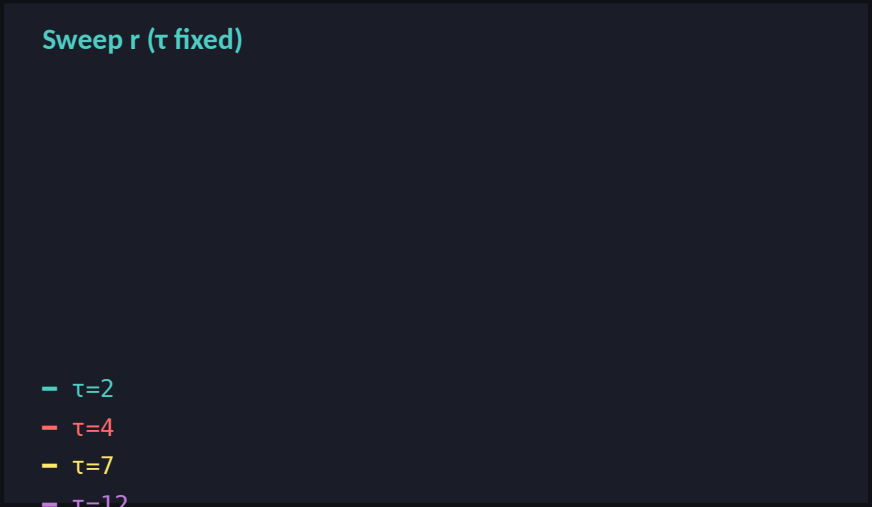


Observe:

- Same parameters → different outcomes each time (stochasticity)
- r,τ distributions visible in histograms
- Low ⟨r⟩ → small outbreak. High ⟨r⟩ → epidemic. Where is the boundary?

MEASUREMENT

What Controls the Epidemic?



Observation: Both r and τ affect the outcome, but neither alone determines it. Different (r, τ) pairs can produce the same epidemic size.

→ **There seems to be a single hidden quantity that combines r and τ.**

What is it?

MEASUREMENT

Transmissibility T

The hidden variable: $T = 1 - (1-r)^{\tau}$

The probability that infection crosses an edge over the entire infectious period.

This single number combines r and τ.

MEASUREMENT

Average Epidemic Size vs. T

Run SIR many times for each T.
Plot average fraction infected vs. T.

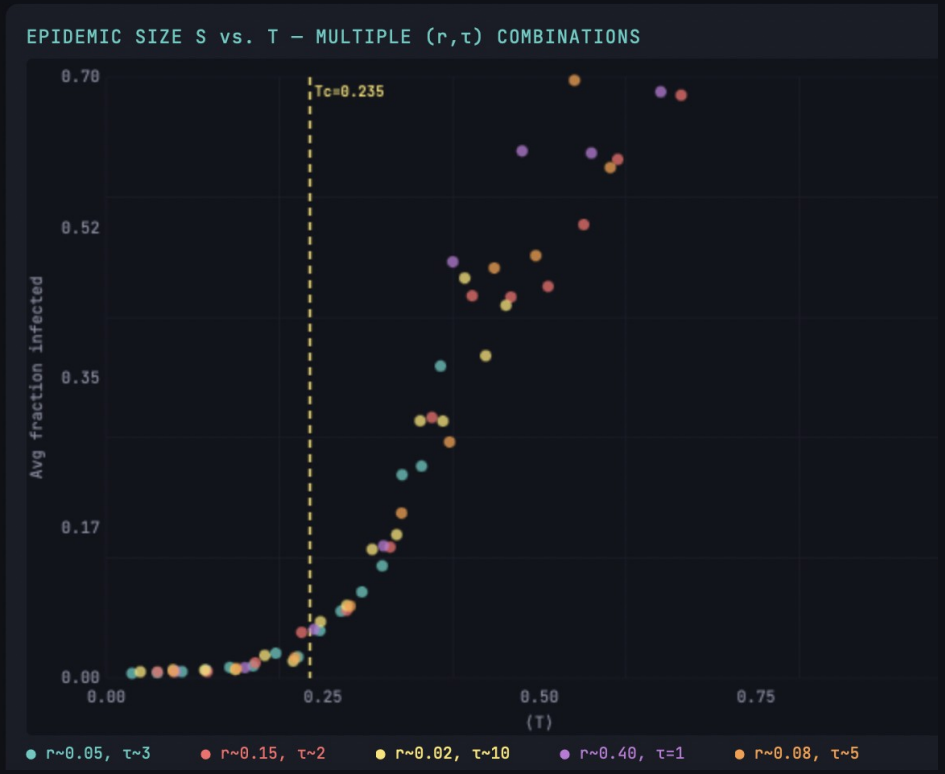
Sharp transition at some Tc.
Below: small outbreaks only.
Above: epidemic possible.

Mean sub-epidemic outbreak size diverges as $T \to T_c^-$

— a phase transition.

Do the Details of r, τ Matter?

Do the Details of r, τ Matter?



5 completely different (r,τ) distributions.

When plotted against ⟨T⟩:

All data collapse onto the same curve.

Individual r,τ variation washes out.
Only the average ⟨T⟩ matters.

Why? → Newman's derivation shows this mathematically.

PART 3 — NEWMAN'S DERIVATION

Step 1: Encoding the Network

Degree distribution → generating function

$$G_0(x) = \sum_k p_k x^k$$

$$G_0'(1) = \langle k \rangle$$

$$G_0''(1) = \langle k(k-1) \rangle$$

Excess degree distribution

Follow a random edge → endpoint has degree $\propto k \cdot p_k$ (size-biased).

Excess degree = degree − 1.

$$G_1(x) = G_0'(x) \,/\, G_0'(1)$$

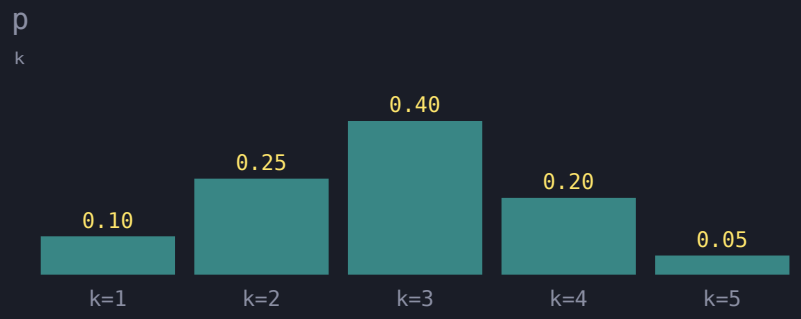
G_0' multiplies by k → size-biasing

$\div \langle k \rangle$ → normalization

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x \frac{G_0'(x)}{G_0'(1)}.$$

Step 1a: What Is $G_0(x)$?

Example: degree distribution of a network



Encode as a polynomial:

$$G_0(x) = 0.10x + 0.25x^2 + 0.40x^3 + 0.20x^4 + 0.05x^5$$

coefficient of $x^k = p_k$

Why a polynomial? Derivatives give moments:

$$G_0'(x) = \Sigma\, k \cdot p_k \cdot x^{k-1}$$

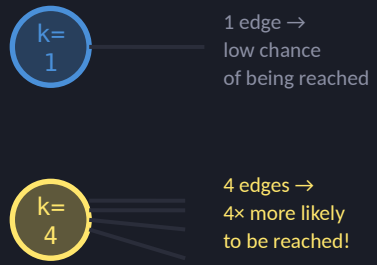
$$G_0'(1) = \Sigma\, k \cdot p_k = \langle k \rangle = 2.75$$

$$G_0''(1) = \Sigma\, k(k-1) \cdot p_k = \langle k(k-1) \rangle = 6.50$$

x is a formal variable — a tool that
lets us extract statistics by differentiation.

Step 1b: Excess Degree $G_1(x)$

Which node do you reach by following a random edge?



Probability of reaching a node $\propto k \cdot p_k$
(not just p_k — this is size-biasing)

Probability of reaching degree-k node:

$$\tilde{q}_k = k \cdot p_k / \Sigma_j\, j \cdot p_j = k \cdot p_k / \langle k \rangle$$

Excess degree = k − 1
(subtract the edge you arrived on)

Generating function for excess degree:

$$G_1(x) = \Sigma_k\, q_k\, x^{k-1}$$
$$= \Sigma_k\, (k \cdot p_k / \langle k \rangle)\, x^{k-1}$$
$$= G_0'(x) / G_0'(1)$$

$G_1'(1) = G_0''(1)/G_0'(1) = \langle k(k-1) \rangle / \langle k \rangle$ — this number will determine Tc.

Step 2: Adding Transmissibility

Node has degree k. Each edge occupied with probability T.
occupied edges ~ Bin(k, T).

Generating function transforms as:

$$G_0(x;\ T) = G_0(1 + (x-1)T)$$

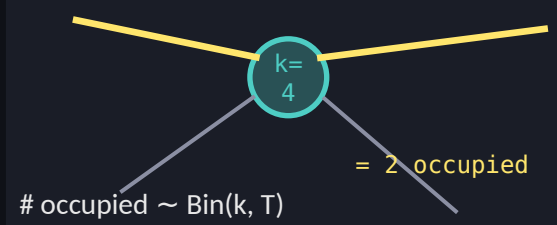
Key insight: When r_{ij} and τ_i are i.i.d. random, the T_{ij} are also i.i.d.
The collective behavior depends only on the average $T = \langle T_{ij} \rangle$.

→ This is why the curves collapsed in the demo.

The math makes it a theorem, not just an observation.

Step 2: Adding Transmissibility

Each edge: occupied (prob T) or not



How G_0 transforms:

An edge is occupied (x) with prob T,
or not occupied (1) with prob 1−T.

GF for one edge's contribution:

$$(1-T)\cdot 1 + T\cdot x = 1 + (x-1)T$$

Node has k edges, each independent:

$$[1 + (x-1)T]^k$$

Average over degree distribution p_k:

$$\Sigma_k\ p_k\ [1+(x-1)T]^k = G_0(1+(x-1)T)$$

Key insight: When r_{ij} and τ_i are i.i.d., the T_{ij} are also i.i.d. → collective behavior depends only on $\langle T \rangle$.

→ This is why the curves collapsed in the demo. The math makes it a theorem.

Step 3: The Recursive Structure

$H_1(x)$ = cluster size distribution reached via a random edge.

The endpoint has excess degree $\sim G_1$.

Each occupied edge leads to a sub-cluster with the same distribution → recursion.

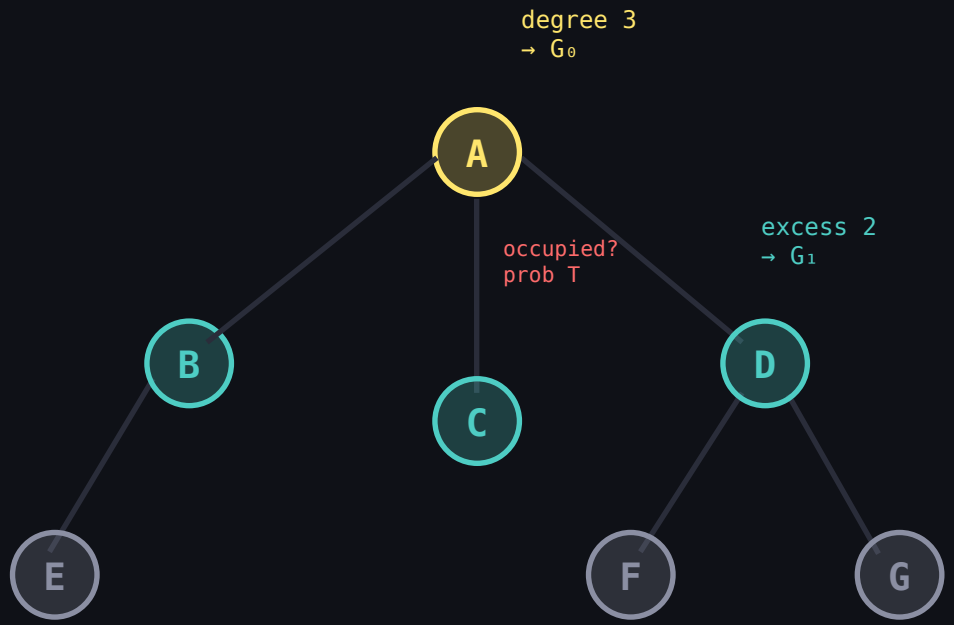
$$H_1(x) = x \cdot G_1(\, H_1(x)\, ;\, T\,)$$

$$H_0(x) = x \cdot G_0(\, H_1(x)\, ;\, T\,)$$

x = this node **$G_1(H_1;T)$** = sub-clusters via occupied edges

This works because the network is **locally tree-like**: no short loops → no correlation between sub-clusters → the recursion is exact.

What G_0, G_1, H_1 Mean on a Network



$G_0(x)$ encodes degree distribution p_k

e.g. this network: $p_1=2/7$, $p_2=2/7$, $p_3=3/7$

$G_0(x) = 2/7 \cdot x + 2/7 \cdot x^2 + 3/7 \cdot x^3$

$G_0'(1) = \langle k \rangle = 15/7 \approx 2.1$

$G_1(x)$ = excess degree distribution

Follow edge A→D: D has degree 3, arrived on 1 edge → excess = 2.

$G_1(x) = G_0'(x)/G_0'(1)$ — derivative does size-biasing (high-k nodes reached more).

H_1 = cluster size via an edge

Follow A→D. If edge is occupied (prob T), D's cluster = D + sub-clusters via F, G. Each sub-cluster is again H_1 → recursion.

$H_1(x) = x \cdot G_1(H_1(x) ; T)$

H_0 = cluster of a random node (A)

A + sub-clusters via B, C, D = $x \cdot G_0(H_1(x);T)$

Step 3→4: From Recursion to ⟨s⟩

We want the mean cluster size ⟨s⟩ = $H_0'(1)$.

Differentiate the self-consistent equation

$$H_1(x) = x \cdot G_1(H_1(x); T)$$

Take d/dx and set x=1. Using chain rule:

$$H_1'(1) = 1 + T \cdot G_1'(1) \cdot H_1'(1)$$

Solve for $H_1'(1)$

$$H_1'(1) = 1 / [1 - T \cdot G_1'(1)]$$

Diverges when denominator → 0

Now $H_0'(1) = \langle s \rangle$

$$H_0(x) = x \cdot G_0(H_1(x); T)$$

$$\langle s \rangle = 1 + T \cdot G_0'(1) \cdot H_1'(1)$$

Substituting:

$$\langle s \rangle = 1 + T \cdot G_0'(1) / [1 - T \cdot G_1'(1)]$$

The denominator controls everything. When it → 0, the mean cluster size diverges → epidemic.

Step 4: Epidemic Threshold

From the previous slide:

$$\langle s \rangle = 1 + T \cdot G_0'(1) / [1 - T \cdot G_1'(1)]$$

Set denominator = 0: $T \cdot G_1'(1) = 1$

Epidemic Threshold

$$T_c = \langle k \rangle / \langle k(k-1) \rangle$$

→ r, τ don't appear — only T and degree distribution

→ Threshold is sharp — genuine phase transition (⟨s⟩ diverges)

→ Network effect: power-law (α=2,κ=10) → Tc=0.329 vs fully-mixed → Tc=0.558

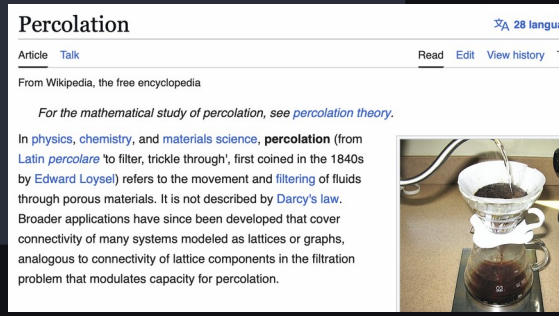
PART 4 — THE BIGGER PICTURE

This Is Percolation

"Each edge occupied with probability T. What is the connected cluster structure?"

= Bond percolation.

Broadbent & Hammersley (1957): will water percolate through a porous rock?


Percolation

28 langu

Article Talk Read Edit View history

From Wikipedia, the free encyclopedia

For the mathematical study of percolation, see percolation theory.

In physics, chemistry, and materials science, **percolation** (from Latin *percolare* 'to filter, trickle through', first coined in the 1840s by Edward Loysel) refers to the movement and filtering of fluids through porous materials. It is not described by Darcy's law. Broader applications have since been developed that cover connectivity of many systems modeled as lattices or graphs, analogous to connectivity of lattice components in the filtration problem that modulates capacity for percolation.

SIR epidemic threshold	↔	Percolation threshold pc
Epidemic (giant component)	↔	Water reaches the other side
Network degree distribution	↔	Lattice / pore structure

Same mathematics → conductivity, gelation, forest fires, network robustness, …

Universality: Same Exponents Everywhere

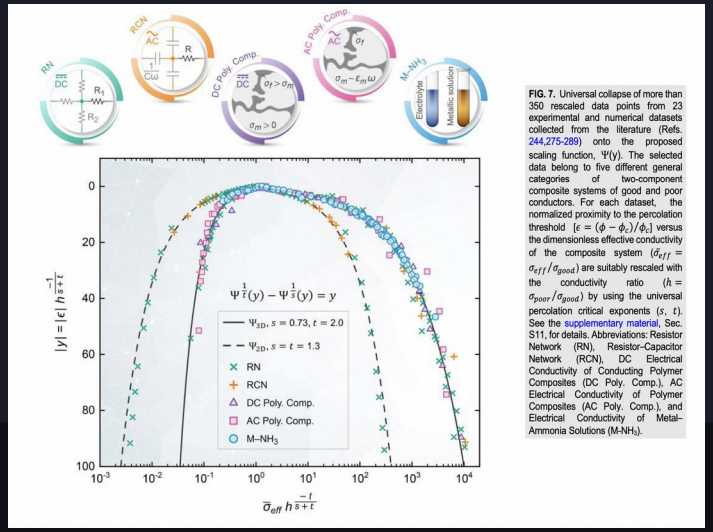


FIG. 7. Universal collapse of more than 350 rescaled data points from 23 experimental and numerical datasets collected from the literature (Refs. 244,275-289) onto the proposed scaling function, $\Psi(y)$. The selected data belong to five different general categories of two-component composite systems of good and poor conductors. For each dataset, the normalized proximity to the percolation threshold $[\epsilon = (\phi - \phi_c)/\phi_c]$ versus the dimensionless effective conductivity of the composite system ($\sigma_{eff} = \sigma_{eff}/\sigma_{good}$) are suitably rescaled with the conductivity ratio ($h = \sigma_{poor}/\sigma_{good}$) by using the universal percolation critical exponents (s, t). See the supplementary material, Sec. S11, for details. Abbreviations: Resistor Network (RN), Resistor-Capacitor Network (RCN), DC Electrical Conductivity of Conducting Polymer Composites (DC Poly. Comp.), AC Electrical Conductivity of Polymer Composites (AC Poly. Comp.), and Electrical Conductivity of Metal–Ammonia Solutions (M-NH_3).

Sarikhani et al. (2022)

23 systems → 1 universal curve

slope = t ≈ 2.0

Threshold ϕc: non-universal (system-specific)

Exponent t: universal (depends only on dimension)

2D: all exponents exact, rational numbers.

Fields Medal 2022 (Duminil-Copin)

https://www.youtube.com/watch?v=da3-dHWb4WU

Renormalization group → near criticality, microscopic details vanish. Only dimension matters.

What the Mathematics Gives Us

1 **Equivalence of seemingly different conditions**

r/τ variation → absorbed into $\langle T \rangle$. Provable, not just observable.

2 **A baseline to measure deviations from**

T_c is exact → deviations tell us what's not captured by the model.

3 **Transfer across phenomena (universality)**

Same math: epidemics ↔ conductivity ↔ gelation ↔ networks.

TAKEAWAY

A Simple Model Is Not a Toy

1. Defined SIR on a random network with per-node τ, per-edge r.

2. Measured: epidemic size vs $T \rightarrow$ sharp threshold; r/τ details don't matter.

3. Derived: $T_c = \langle k \rangle / \langle k(k-1) \rangle$ via generating functions + self-consistent equations.

4. Connected: this is percolation $\rightarrow$ universality across physics, biology, materials.