Degree Based Attacks and Defense Strategies in Complex Networks

Prof. Reuven Cohen and Aviv Yehezkel

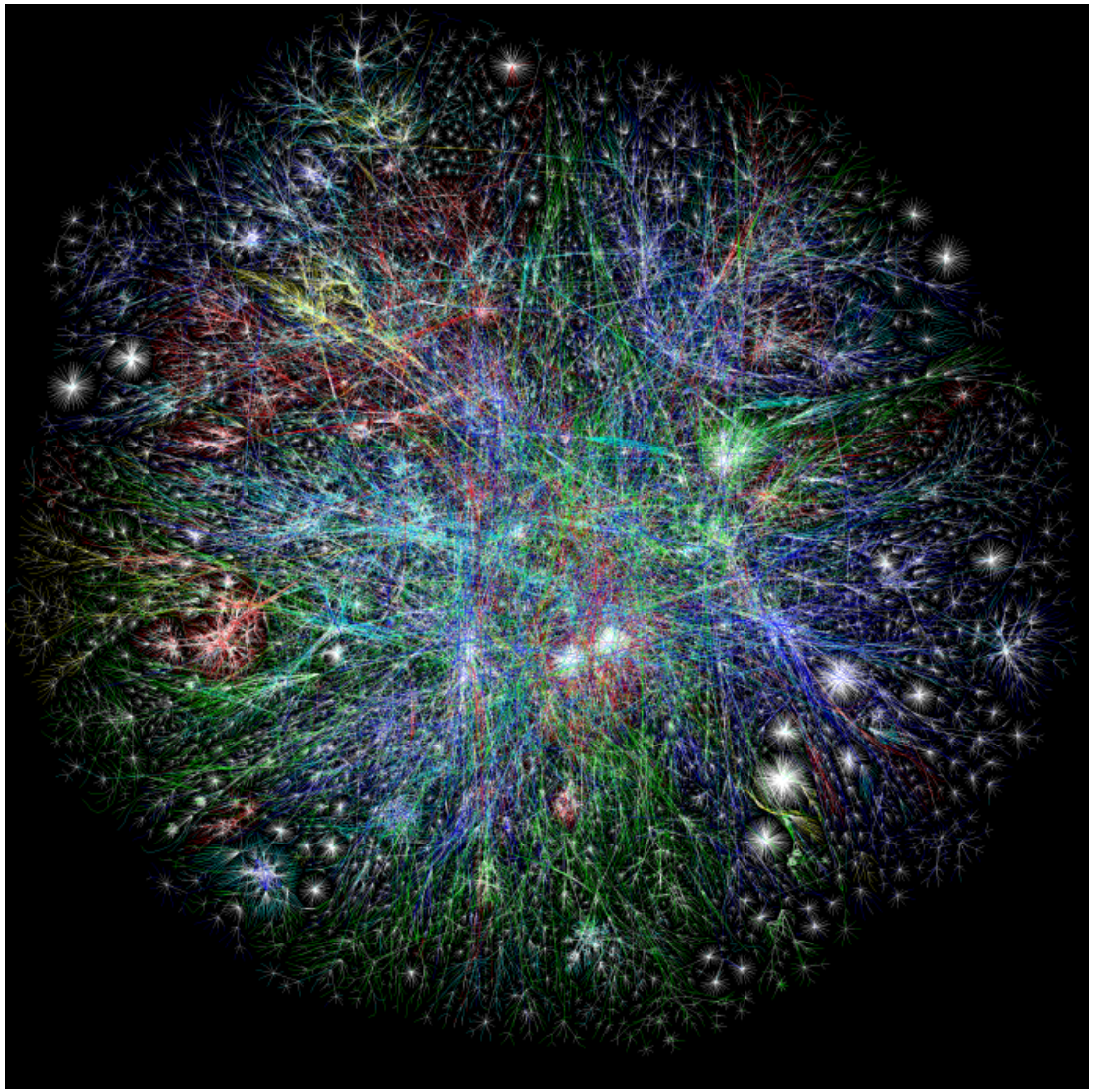
Applied Math Seminar

25 November 2012

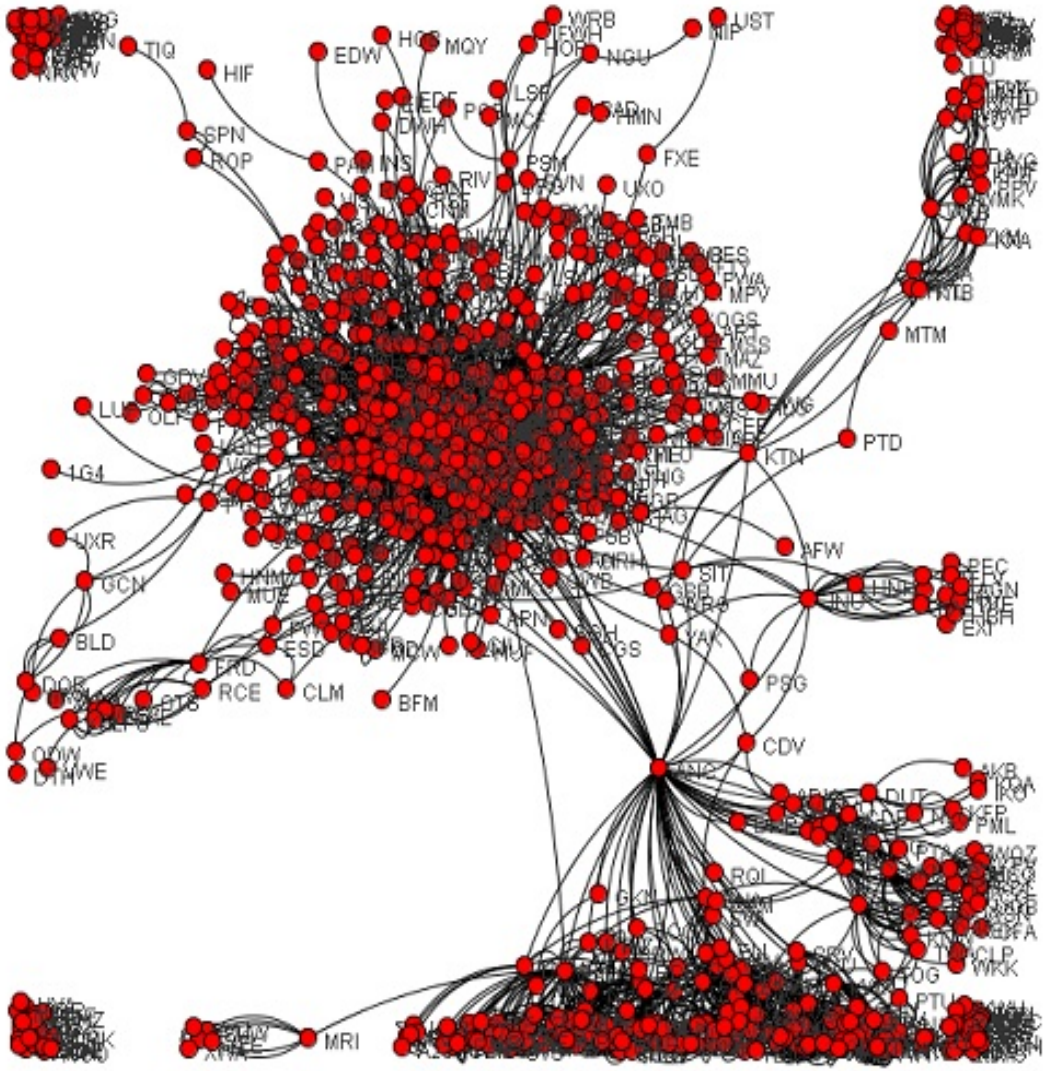
Introduction to Complex Networks

A world full of networks:

- Communication networks
- The world-wide-web Internet
- The electrical power grid
- Airline networks
- Social networks



Internet routers network



US airport network

Friendship network

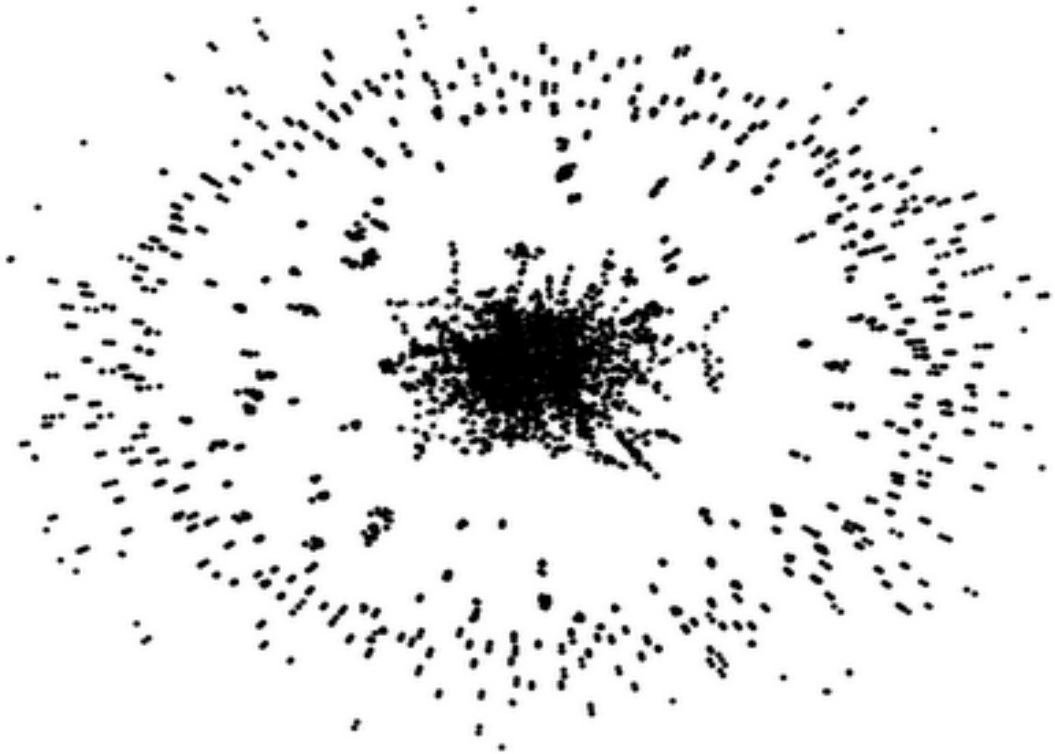
The Beginning of Complex Networks

- Graphs, describing mathematical concepts in networks, usually focus on the properties of special graphs
- 1960s - Paul Erdős and Alfred Rényi (ER)- the random graph theory new concept
- End of 20th century - The classical random graphs theory fails to describe many real world networks
- The beginning of Complex Networks
- The Scale-Free Character

$$P(k) \sim k^{-\gamma}, \quad k = m, ..., K_{max}$$

Complex Networks - The Basics

- Connected component - group of nodes connected internally, but disconnected from other components
- Giant connected component - connected component with size (number of nodes) proportional to that of the entire network



Authorship network consists of one giant connected component with 4600 authors (60.3%), along with a large number of much smaller components (the second largest component has 103 authors).

Attacks in Networks

- Network's robustness
- The Standard approach - the removal of nodes influences the structure of the network
- The network will keep its ability to perform its tasks as long as a giant connected component exists
- Percolation threshold p_c - critical fraction of removed nodes needed for destroying the network
  For $p < p_c$, a giant connected component still exists
  For $p > p_c$, only small connected components appear

Random and Intentional Attack

- Random removal
- Complex networks show great stability even against large number of random removals

- Intentional attack
- The first node which is removed has the highest degree. Then the node with the second highest degree is removed, and so on
- It is usually enough to remove a very small fraction of the highly connected nodes to completely destroy the network

Intentional Attack - Why Not?

- Requires complete knowledge of the network structure
- Complete knowledge is either not available, does not exist or is not accessible to the attacker
- The goal is to find other attack strategies that does not rely on full knowledge of the network topology
- Degree Based Attacks

Degree Based Attacks

- $\pi(k_i) =$ the probability that a node with degree k_i will be removed:

$$\pi(k_i) = \frac{\alpha k_i}{K_{max}}, \quad 0 < \alpha \leq 1 \tag{1}$$

- α can represent the level of vulnerability of nodes, or the level of knowledge of the node degrees by the attacker

Numerical Treatment

- Build the network for the exponent γ, the size of the network N, the minimal degree m
- Go over the network and remove each node with probability $\pi(k_i)$
- Calculate the two moments to get $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}$
- If $\kappa \geq 2$, then a giant component still exist in the network

Analytical Treatment

- K_i = maximal degree of the network at the ith scan
- After d scans the probability that a node of initial degree k will still be functional is:

$$\rho(k,d) = \prod_{i=1}^{d} \left(1 - \frac{\alpha k}{K_i}\right) . \tag{2}$$

- d_c = number of scans
- In the last scan only a fraction c of the network was scanned until it was destroyed and the process ended

Analytical Treatment - The Solution (1)

The probability that a node of degree k will still be functional at the end of this process:

$$\begin{aligned}\rho(k) &= c\prod_{i=1}^{d_c}\left(1-\frac{\alpha k}{K_i}\right)+(1-c)\prod_{i=1}^{d_c-1}\left(1-\frac{\alpha k}{K_i}\right)\\ &= c\rho(k,d_c)+(1-c)\rho(k,d_c-1)\,. \qquad (3)\end{aligned}$$

Analytical Treatment - The Solution (2)

The condition for the existence of a giant component after d scans is as follows:

$$\sum_{k=m}^{K} \frac{P(k)k(k-1)}{\langle k \rangle} \rho(k,d) = 1 \,. \tag{4}$$

Finally, the critical threshold of removed nodes:

$$p_c = \sum_{k=m}^{K} P(k)(1-\rho(k)) \,. \tag{5}$$

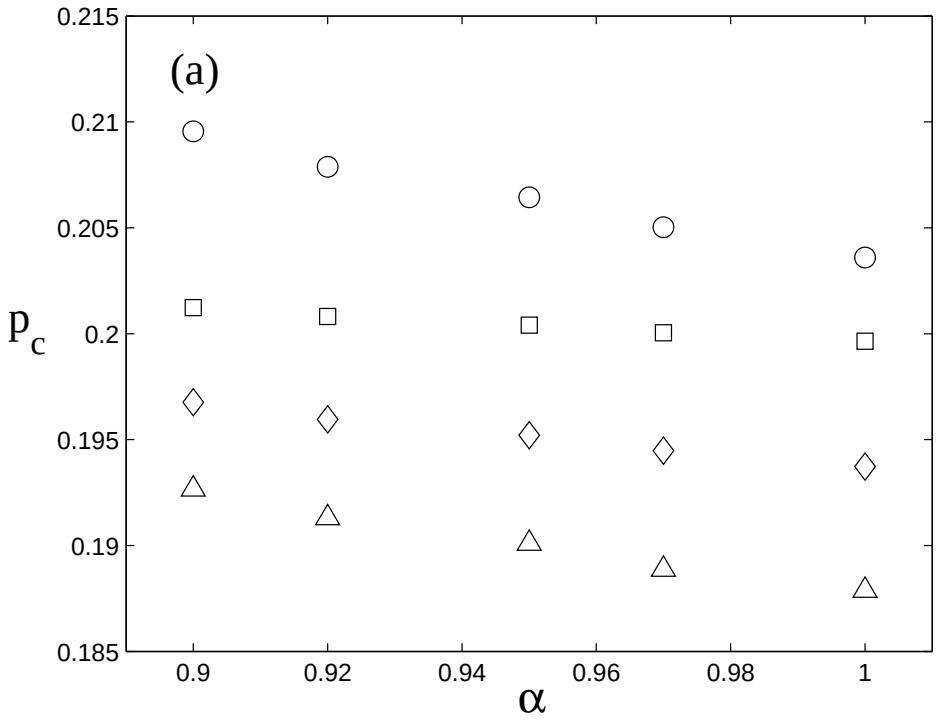


Figure: (1a) Values of p_c vs α for different γ values: $\gamma = 2.5$ ($\circ$), $\gamma = 2.3$ ($\square$), $\gamma = 2.7$ ($\diamond$) and $\gamma = 2.1$ ($\triangle$)

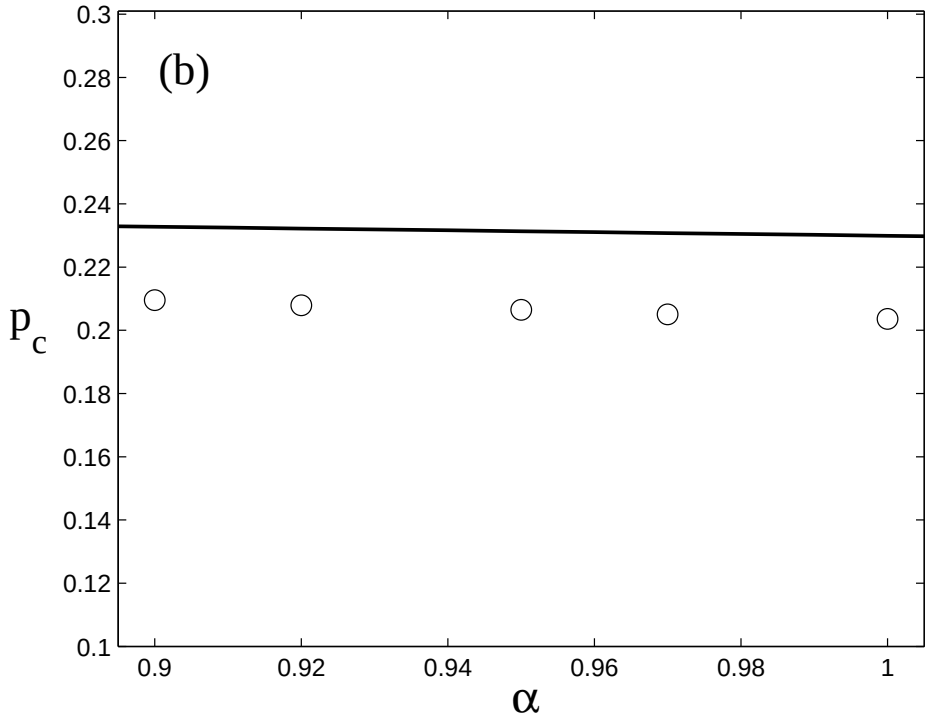


Figure: (1b) Values of p_c vs α for $\gamma = 2.5$. Symbols represent simulation data and solid line is the theoretical prediction

Degree Based Attacks - The Conclusion

A very small fraction of the network has to be removed in order to destroy the network, even without full knowledge of the network topology and connectivity

Defense Strategies

- What is the best way to defend against such attacks?
- Different defense strategies
  Adding links to different groups of nodes - low degree, mid degree, high degree and very high degree nodes
- Which is the most efficient defense strategy?
- Which strategy leads to the highest network robustness improvement?

Defense Strategies

- $C(i,j)$ = a "cut of the network" that include all the nodes with initial degree $i \leq k \leq j$
- Adding E new links between nodes from $C(i,j)$
  The addition is done by randomly selecting pairs of nodes from the cut and linking them

Analytical Treatment in Defense

$P(i,j) = \sum_{k=i}^{j} P(k)$ is the proportional size of the cut $C(i,j)$ in the network.

The probability that a node from $C(i,j)$ will be chosen and get an additional link is $\frac{2E}{P(i,j)N}$.

Therefore, after d scans the probability that a node from $C(i,j)$ will still be functional is now:

$$\rho(k,d) = \frac{2E\rho(k+1,d)}{P(i,j)N} + \left(1 - \frac{2E}{P(i,j)N}\right)\rho(k,d)\,. \qquad (6)$$

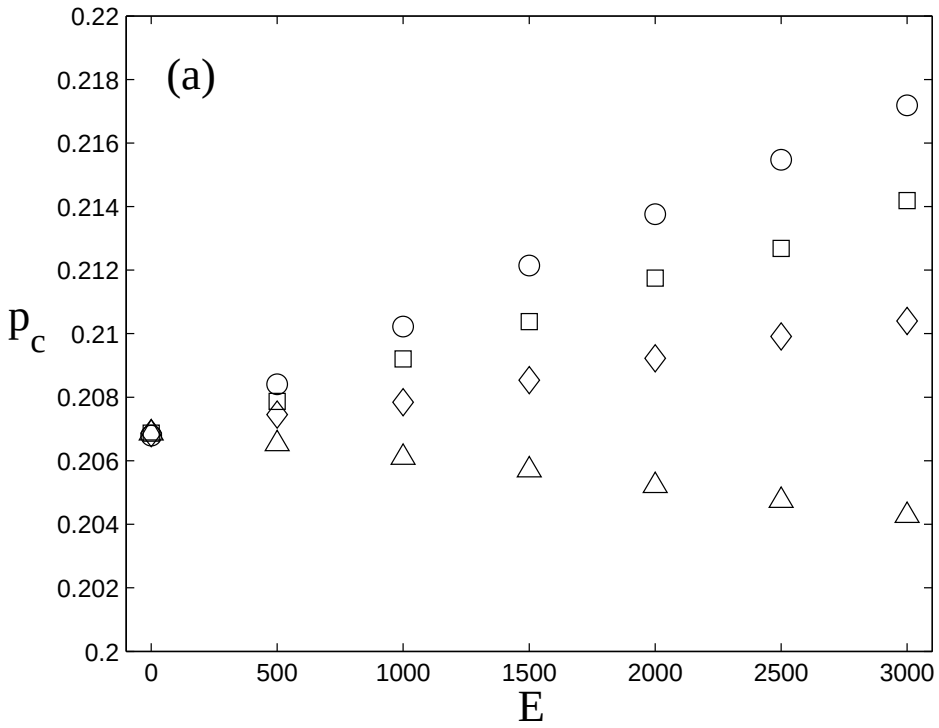


Figure: (2a) Values of p_c vs E for different defense strategies according to simulation data. The symbols represent low degree ($1 \leq k \leq 2$, $\square$), mid degree ($3 \leq k \leq 6$, $\circ$), high degree ($7 \leq k \leq 10$, $\diamond$) and very high degree ($10 \leq k \leq K_{max}$, $\triangle$) defense strategies

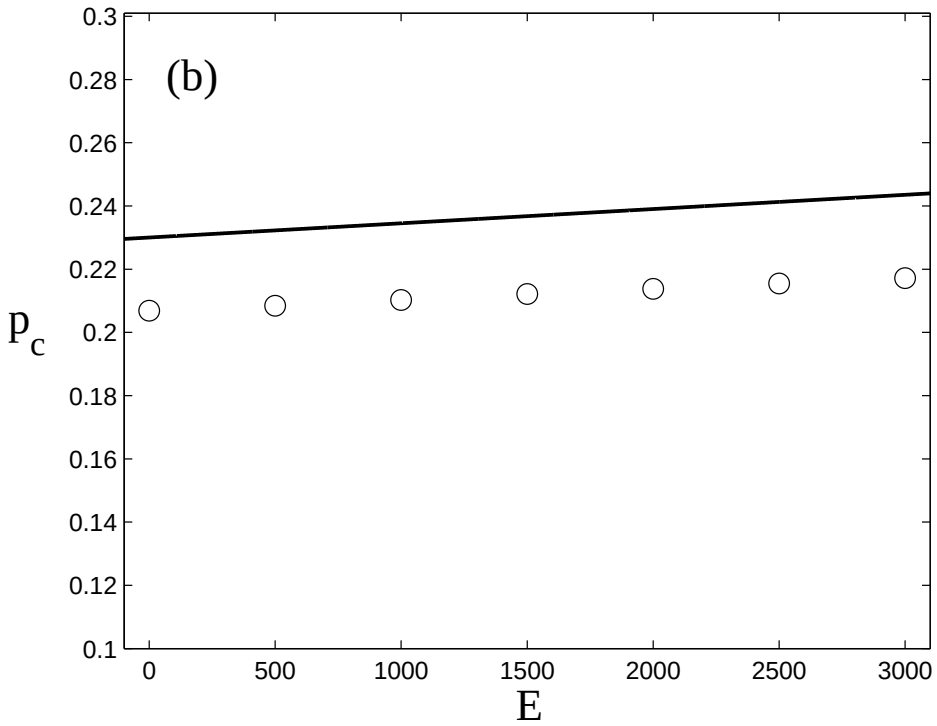


Figure: (2b) Values of p_c vs E for mid degree nodes defense strategy for nodes with initial degree $3 \leq k \leq 6$. Symbols represent simulation data and solid line is the theoretical prediction

Defense Strategies - The Conclusions

- p_c of the mid degree defense is higher than the same p_c for low and high degree defenses
- By defending the mid degree nodes, the network becomes more robust to degree based attacks
- For very high degree node defense, the network robustness actually decreases with the added edges

- **The mid degree nodes are very important to the robustness of the network. New links insertion makes the network much more resilient to degree based attacks**

Defense Strategy Verification

Checking the model for different attacks

- $\pi(k) = \alpha \left(k/K_{max}\right)^2$
- $\pi(k) = \alpha \sqrt{k/K_{max}}$
- In both cases the mid degree nodes defense proves again to be the most efficient

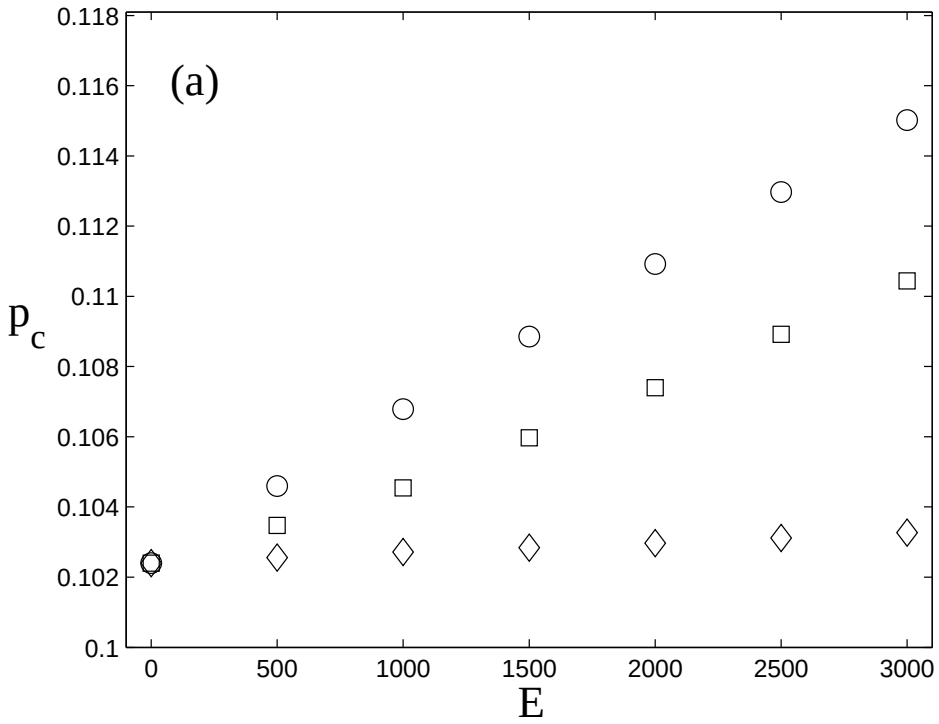


Figure: (3a) Values of p_c vs E for different defense strategies according to simulation data. The symbols represent low degree ($1 \leq k \leq 2$, $\square$), mid degree ($3 \leq k \leq 6$, $\circ$) and high degree ($7 \leq k \leq 10$, $\diamond$) defense strategies for degree based attack with $\pi(k) = \alpha \left(k/K_{max}\right)^2$

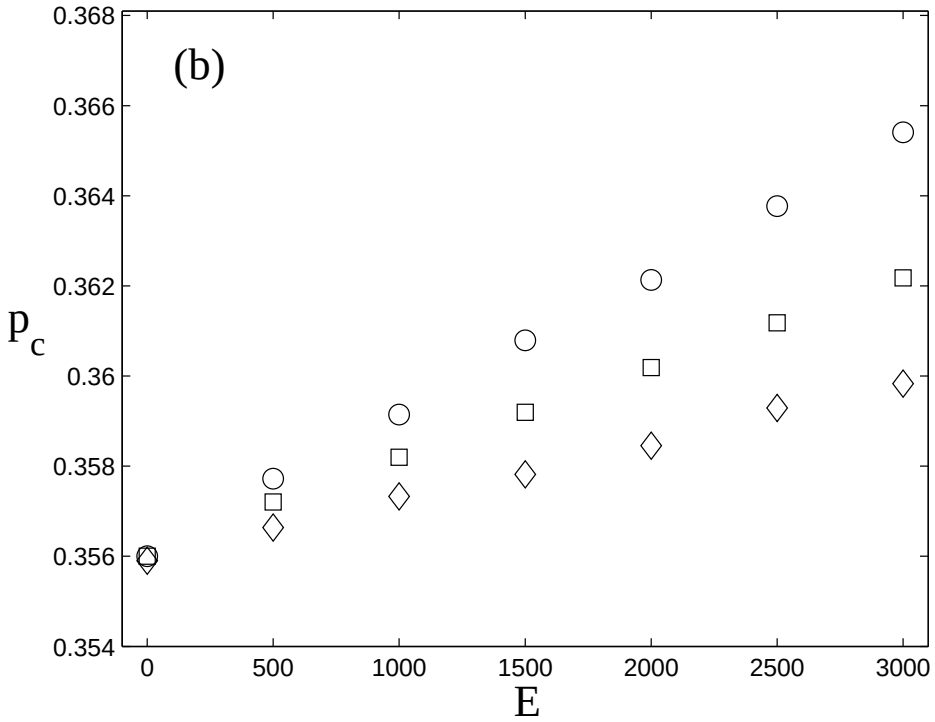


Figure: (3b) same as (a) for degree based attack with $\pi(k) = \alpha\sqrt{k/K_{max}}$

Scenario - Attacks and Defenses in the Internet

- Simulation of the Internet using real data according to the DIMES project
- Initial network - measured map of the Internet's AS level autonomous systems
- Effect study - different defense strategies

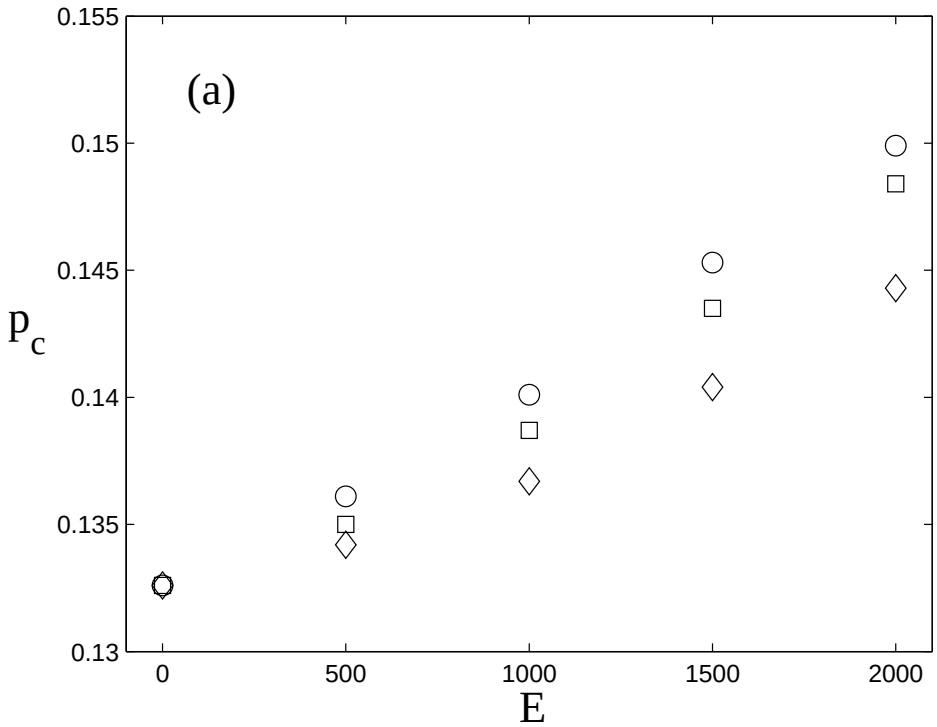


Figure: (4a) Values of p_c vs E according to simulation data based on AS graphs of the Internet from the past years for degree based attack with $\pi(k) = \alpha\,(k/K_{max})$. The attack parameter is $\alpha = 1$

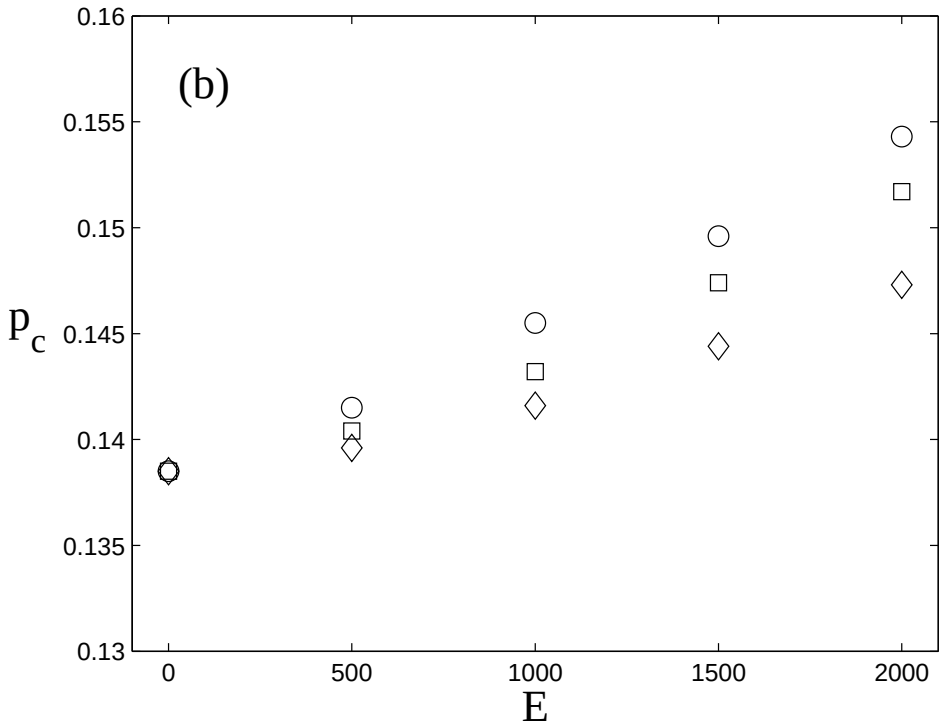


Figure: (4b) same as (a) with the attack parameter $\alpha = 0.7$

Internet Scenario - The Conclusions

- The figures verify the previous results
- The p_c of mid degree nodes defense is higher than the same p_c for low and high degree nodes defenses

- **This result verifies the previous observation that defending the mid degree nodes leads to the highest improvement of network robustness against degree based attacks**

Conclusions

- Partial knowledge of the network structure is sufficient to destroy the network by removing a very small fraction of the nodes
- The Internet can be destroyed completely with very little effort even without full information about its topology
- Defending the mid degree nodes makes the network more robust to degree based attacks