Yandex Research

# Challenges of Generating Structurally Diverse Graphs



Fedor Velikonivtsev

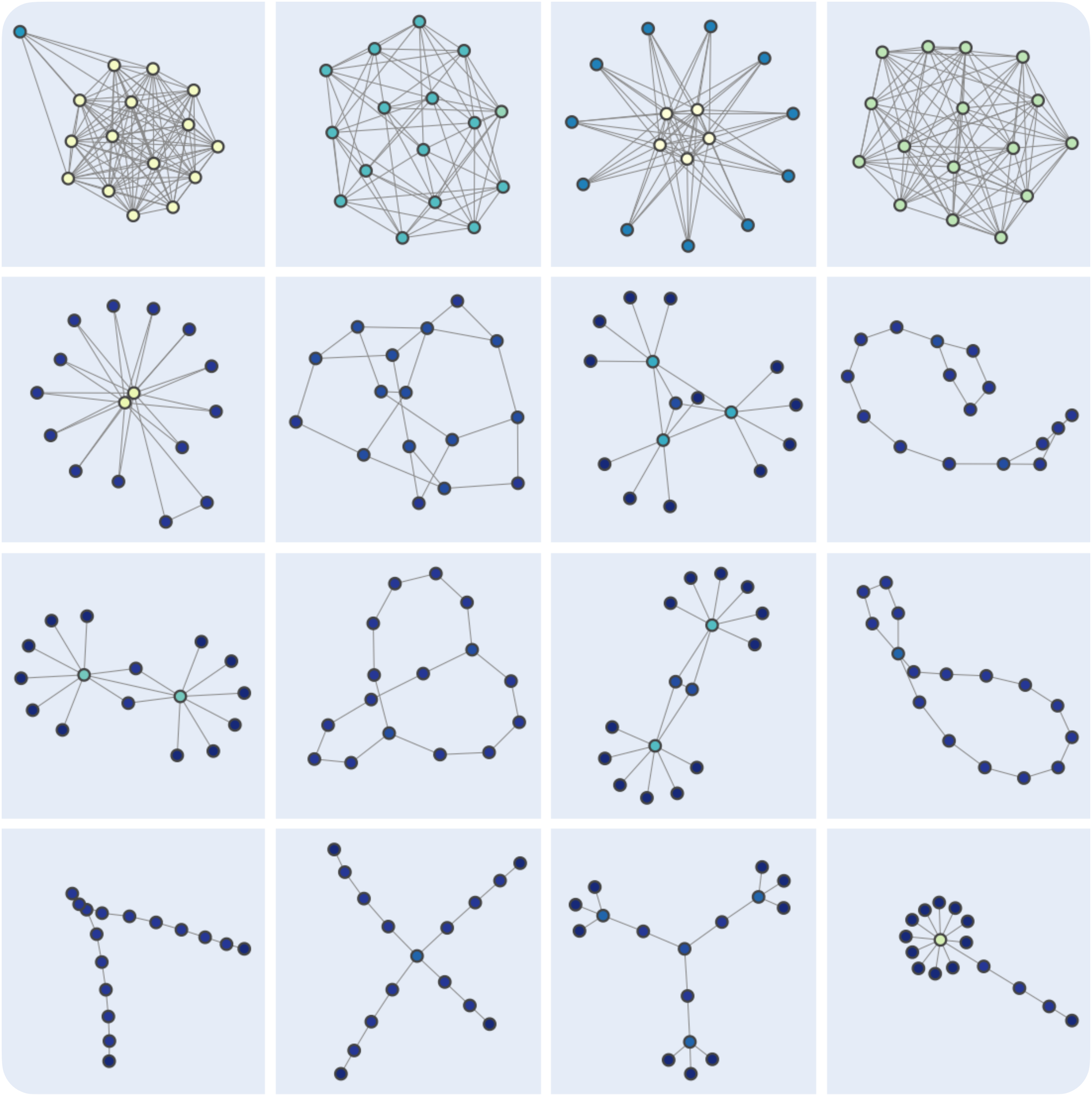

Mikhail Mironov


Liudmila Prokhorenkova

# Motivation

How to use diverse graphs?

- Analyze graph algorithms
- Evaluate heuristic/neural approximations of graph algorithms
- Evaluate GNNs and their expressive power
- Use as training data



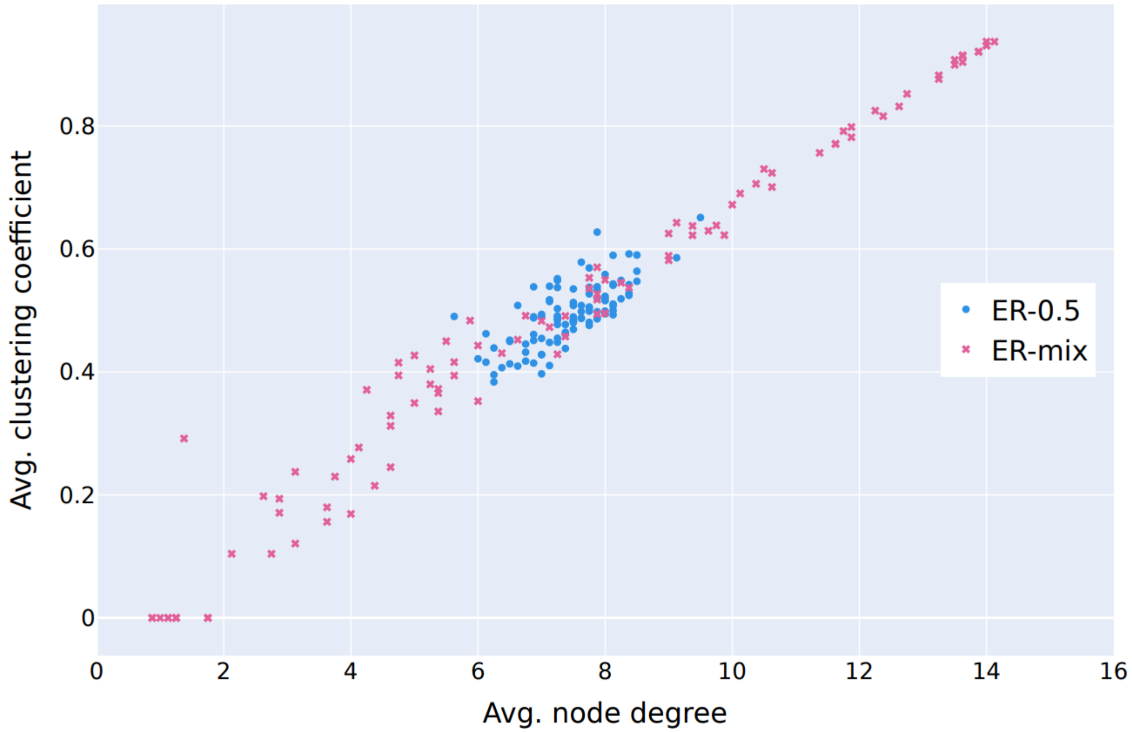
# Problem setup

What we want:

- Generate graphs with diverse structural properties

Why not sample graphs uniformly at random?

- They are similar



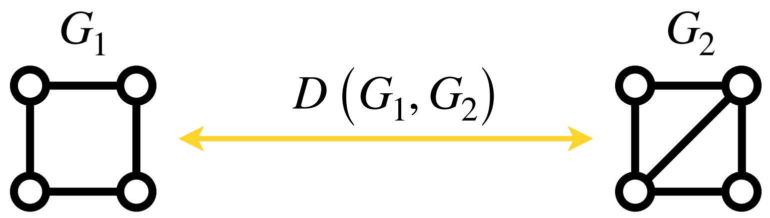

Joint distribution of graph characteristics of Erdős–Rényi models with parameter $p = 0.5$ (ER-0.5) and with uniformly-chosen $p$ (ER-mix)

# How to measure diversity?

- We are given $N$ graphs $\{G_1, \ldots, G_N\}$
- Each graph $G_i$ has $n$ nodes

# How to measure diversity?

- We are given $N$ graphs $\{G_1, \ldots, G_N\}$
- Each graph $G_i$ has $n$ nodes




How we measure diversity:

- Choose some graph distance $D\left(G_i, G_j\right)$

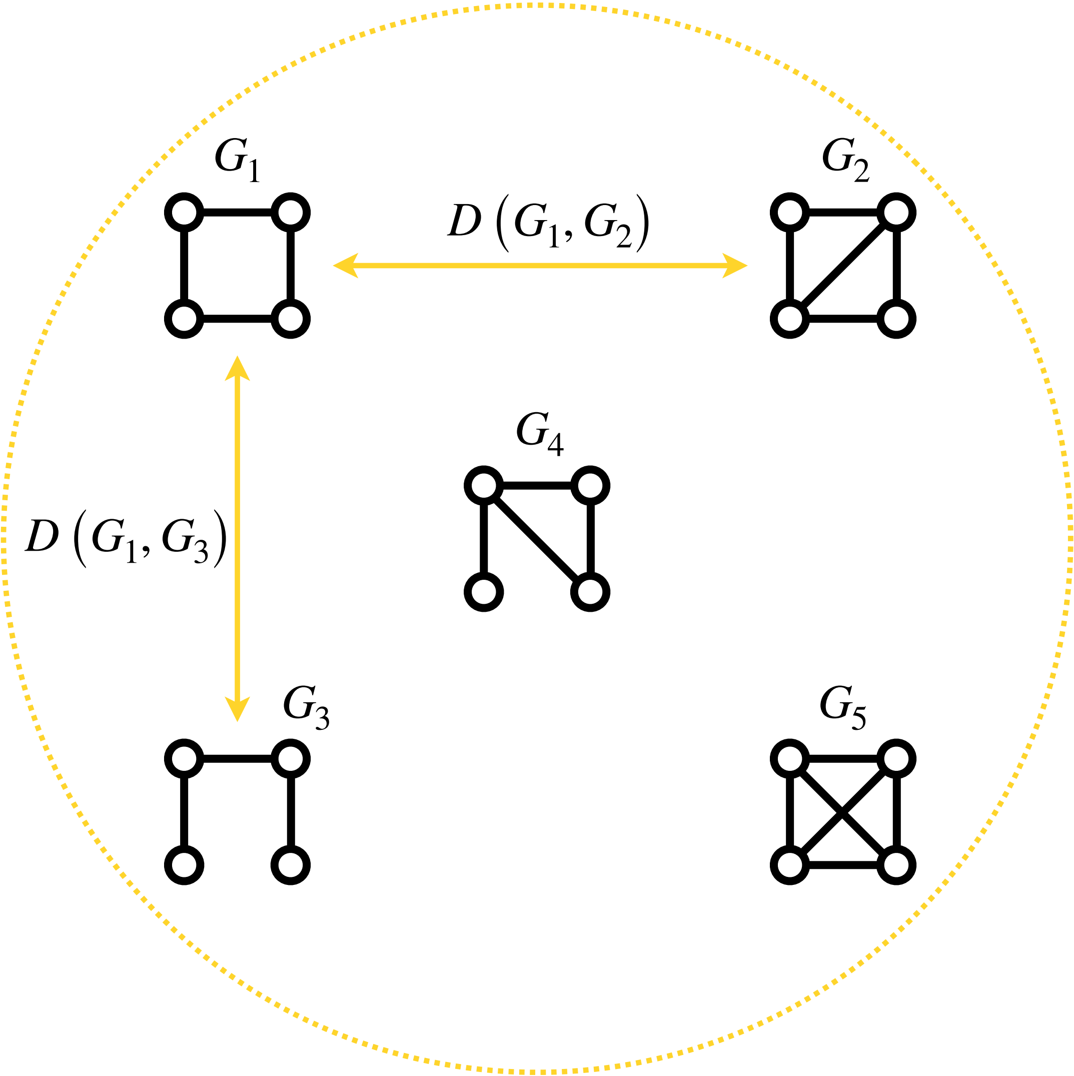
# How to measure diversity?

- We are given $N$ graphs $\{G_1, \ldots, G_N\}$
- Each graph $G_i$ has $n$ nodes

How we measure diversity:

- Choose some graph distance $D\left(G_i, G_j\right)$
- Define diversity as a function of all pairwise distances
- We use **Energy** as a measure of diversity:

$$-\frac{2}{N(N-1)} \sum_{i<j} \frac{1}{D(G_i, G_j)}$$

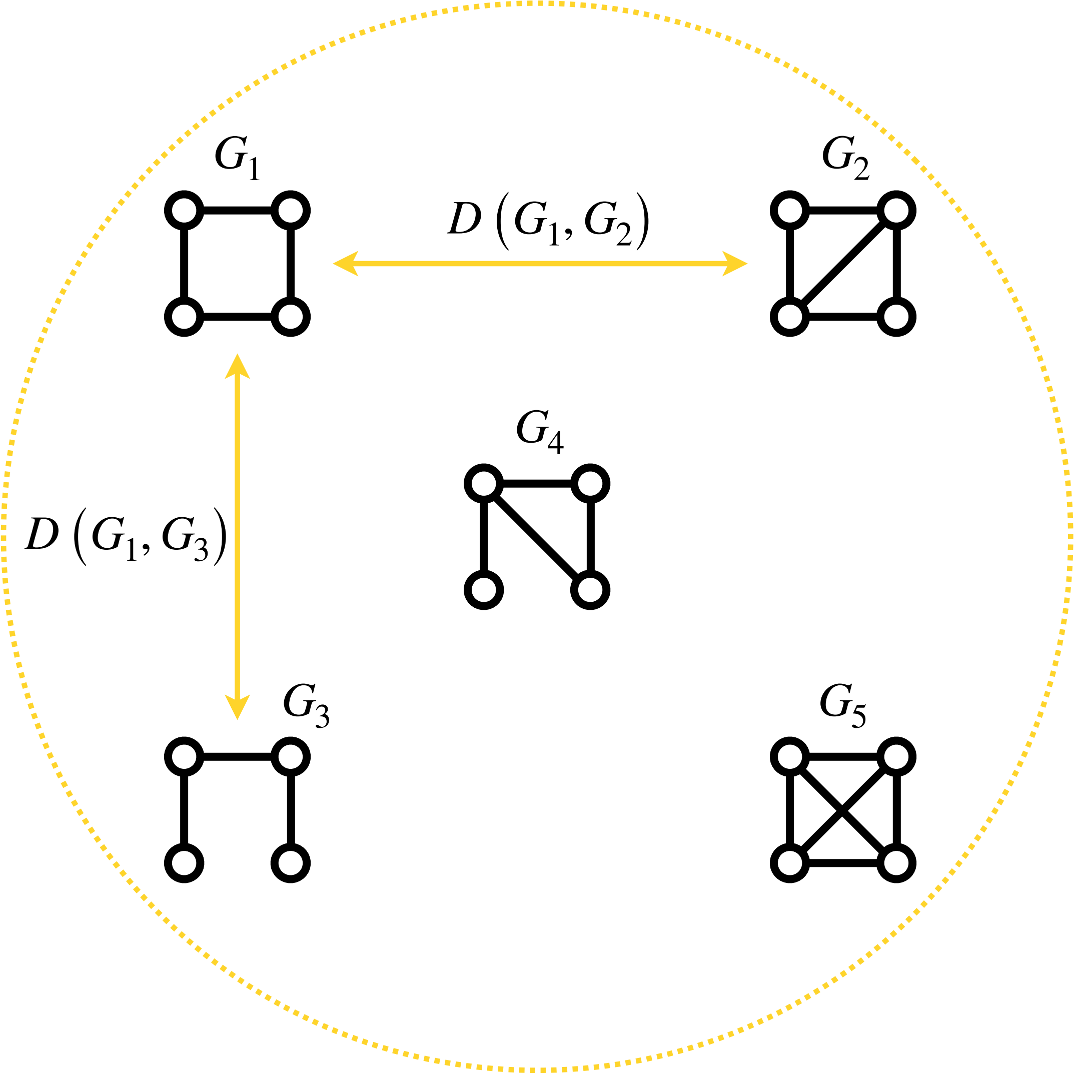


# How to measure diversity?

- We are given $N$ graphs $\{G_1, \ldots, G_N\}$
- Each graph $G_i$ has $n$ nodes

How we measure diversity:

- Choose some graph distance $D\left(G_i, G_j\right)$
- Define diversity as a function of all pairwise distances
- We use **Energy** as a measure of diversity:
$$-\frac{2}{N(N-1)}\sum_{i<j}\frac{1}{D(G_i, G_j)}$$
- Why not, e.g., average distance? See the paper



# Algorithms for optimizing diversity

# Algorithms for optimizing diversity

- Greedy algorithm

## Greedy algorithm

- Uses a large pre-generated set of different graphs, e.g., from random graph models
- Greedily selects graphs from this set to optimize diversity

# Algorithms for optimizing diversity

- Greedy algorithm
- Genetic algorithm

## Genetic algorithm

- Define crossover and mutation operations
- Select parent graphs and generate a child graph that inherits structural traits from its parents
- Update population with a new graph if this increases diversity
- Iterate the procedure

# Algorithms for optimizing diversity

- Greedy algorithm
- Genetic algorithm
- Local optimization (LocalOpt)

## LocalOpt algorithm

- Start with a sufficiently diverse set
- Iteratively modify each graph by making small random changes
- Accept the change if diversity increases
- Iterate the procedure

# Algorithms for optimizing diversity

- Greedy algorithm
- Genetic algorithm
- Local optimization (LocalOpt)
- Iterative graph generative modeling (IGGM)

## IGGM algorithm

- Train a generative model on a set of random graphs
- Use the model to produce a large set of candidate graphs
- Select a diverse subset from the generated set (e.g., via Greedy)
- Train a generative model on this more diverse set
- Repeat the procedure

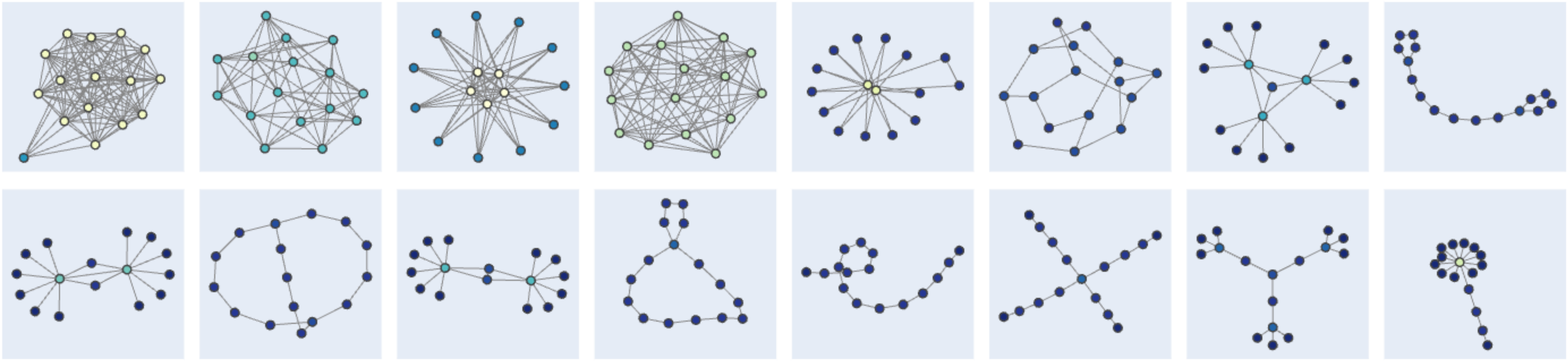
# Algorithms for optimizing diversity

- Greedy algorithm
- Genetic algorithm
- Local optimization (LocalOpt)
- Iterative graph generative modeling (IGGM)

The above approaches can be combined in different ways

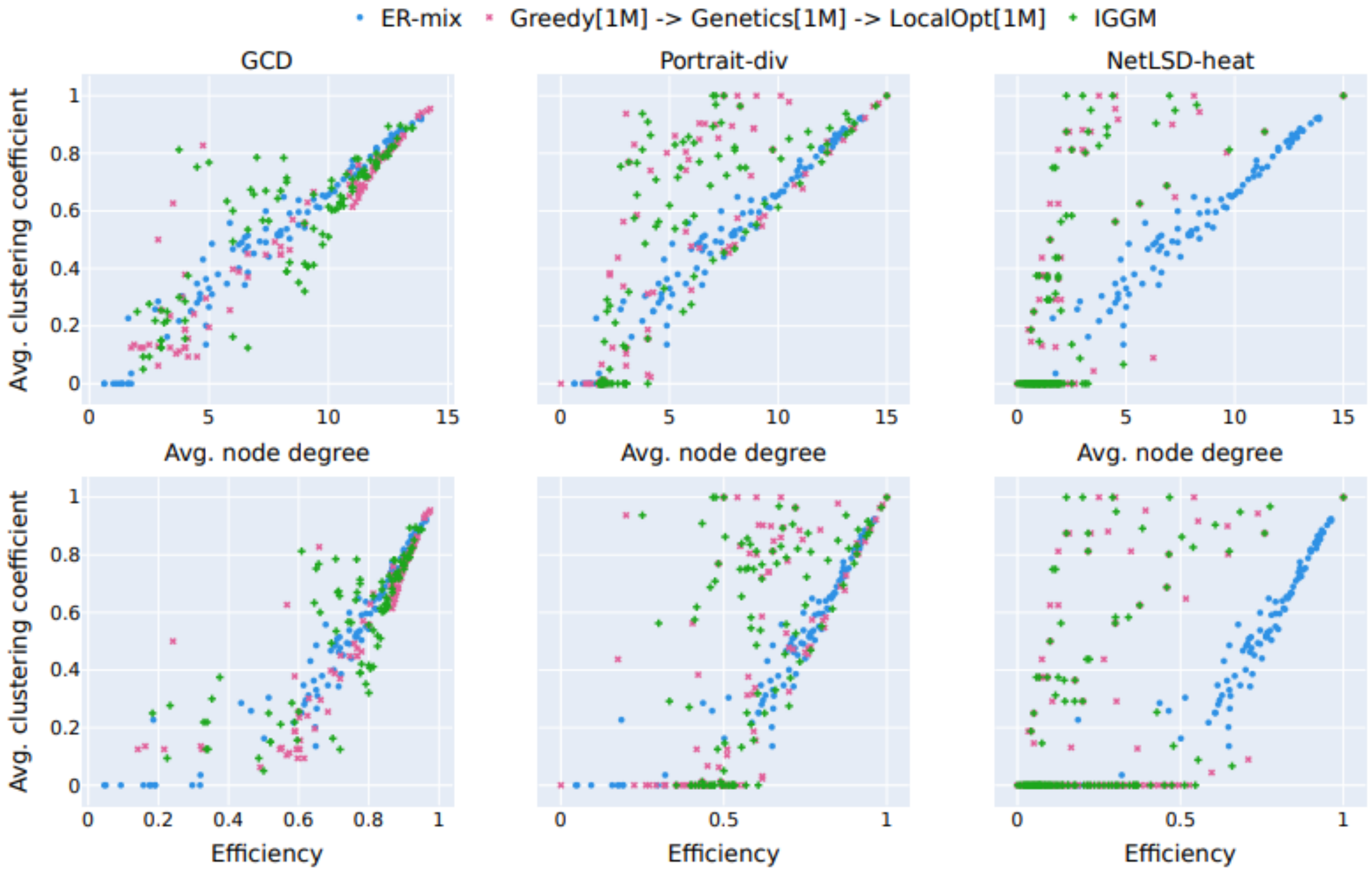
# Results

- Generated graphs have various structural patterns:



# Results

- Graphs have more diverse combinations of characteristics than random graphs:



# Open challenges

- Analysis of diversity measures
- Developing more scalable approaches
- Developing more sophisticated algorithms
- Conditioning on specific properties of graphs
- Applying diverse graphs in practice

Yandex Research

# Thank you!