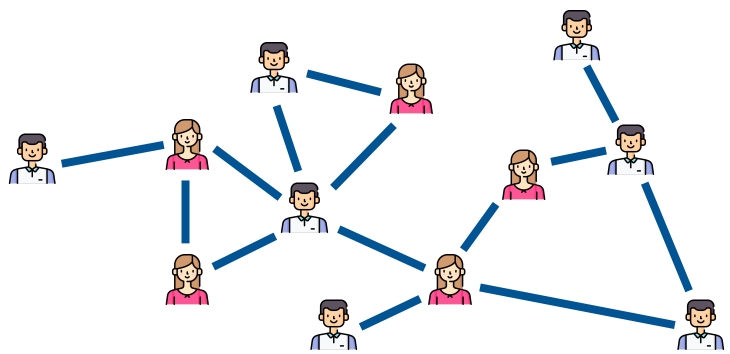
Bayesian optimisation of graph-based functions

Xiaowen Dong

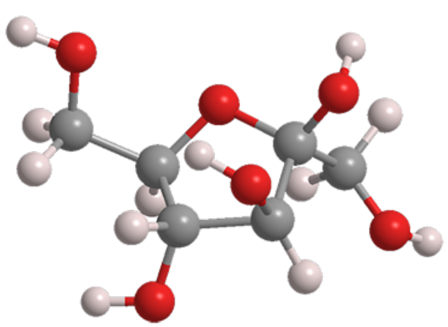
Department of Engineering Science
University of Oxford

Graph structures are pervasive

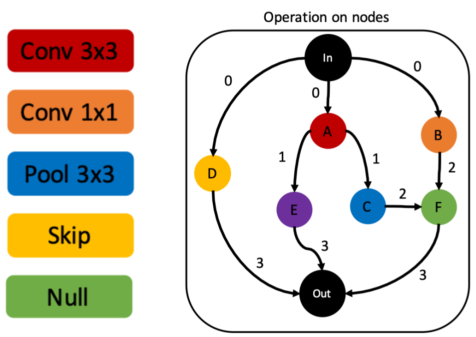


social network

transportation infrastructure



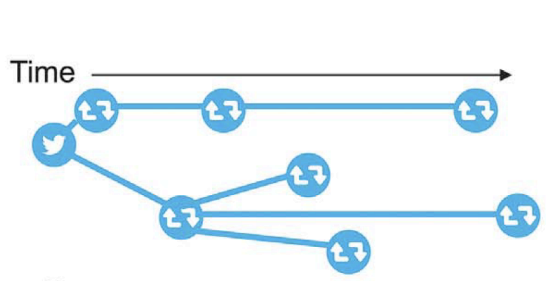
chemical structure



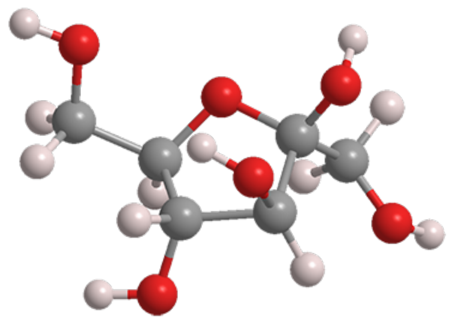
computational architecture

Graph-based functions are pervasive

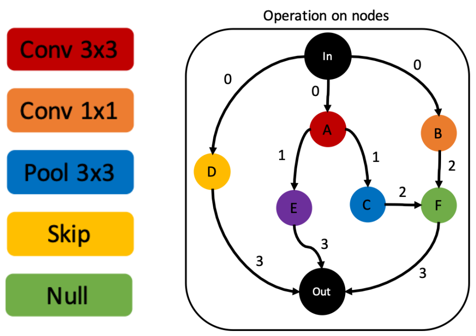
- Graph-wise functions



function: authenticity of news in original post
task: fake news detection



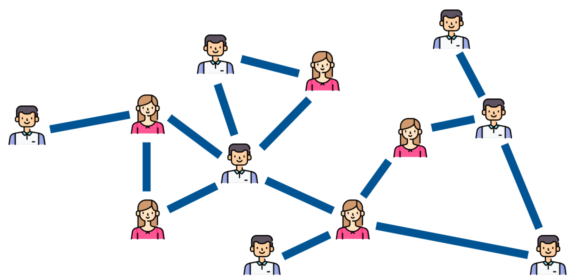
function: potential of molecule inhibiting bacteria
task: drug discovery



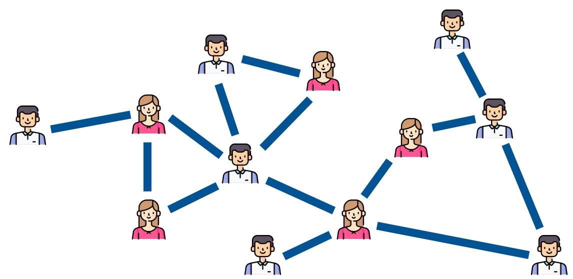
function: test performance of cell-based architecture
task: architecture search

Graph-based functions are pervasive

- Node-wise functions



function: influencing power of individuals
task: influence maximisation



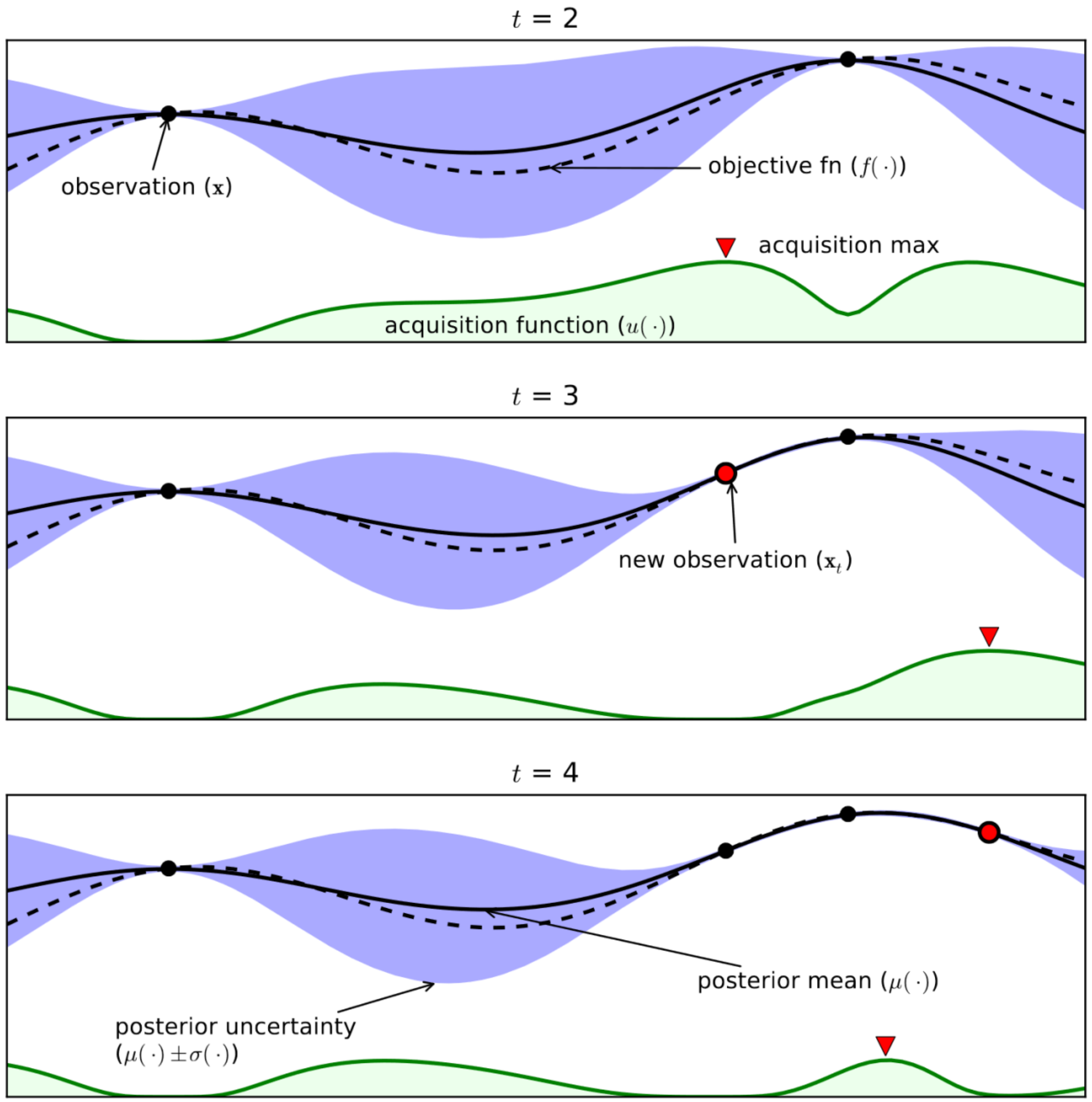
function: infection time of individuals
task: finding patient zero

function: criticality of road junctions
task: resilience monitoring

Optimisation of graph-based functions

- Diverse problem formulations
  - graph-wise functions
  - node-wise functions
- Search space is challenging
  - discrete and combinatorial search space
  - large number of candidates
  - graph may not be known beforehand (e.g., epidemiological contact network, offline social network)
- Functions are often **black-box** and **expensive to evaluate**
  - no analytical form or gradient information
  - often requires simulation or real-world intervention

Bayesian optimisation



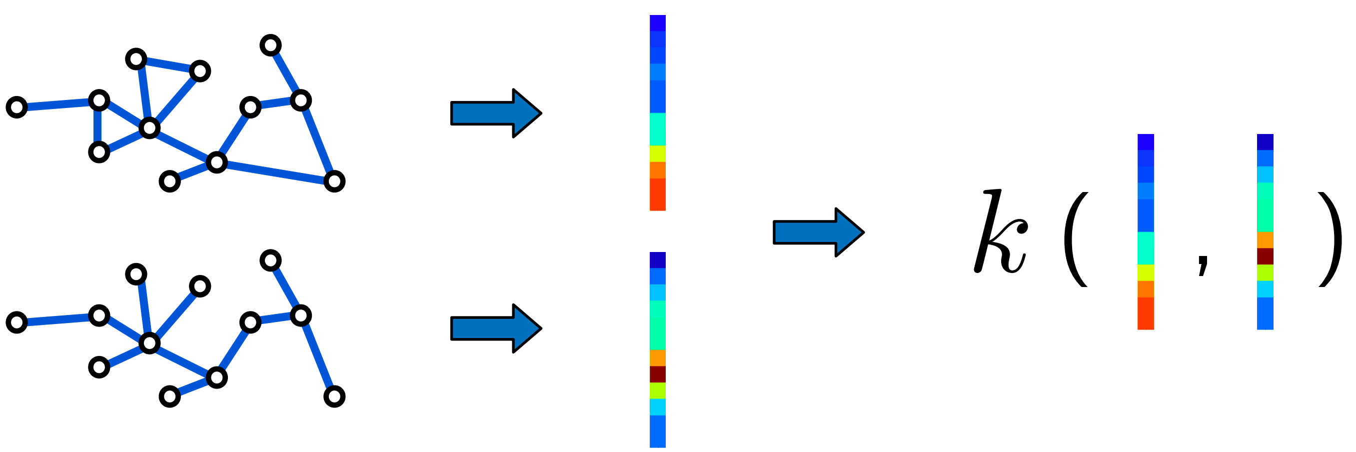
1. query function value at currently chosen point
2. fit **surrogate model**, e.g., Gaussian processes (GPs):

   $f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$
3. use **acquisition function** to find next query point, e.g., upper confidence bound (UCB):

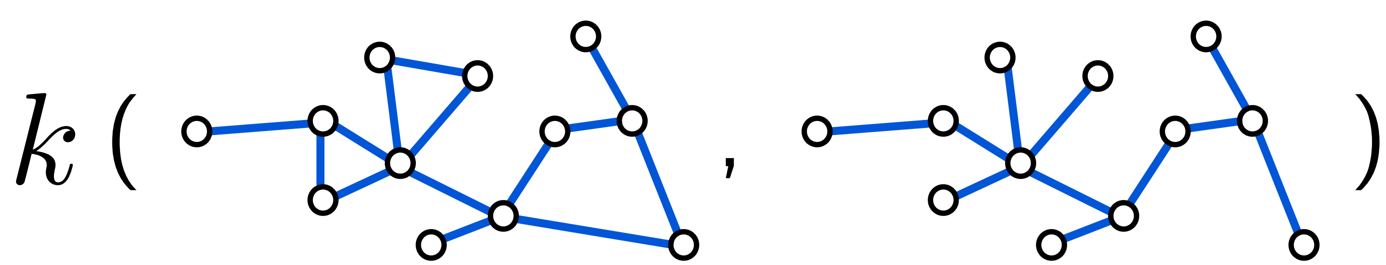
   $a_{\text{UCB}}(x; \beta) = u(x) + \beta\sigma(x)$
4. repeat steps 1-3 until query budget exhausts

Kernels for graph-based functions

- Graph-wise functions
  - key is comparison **between two graphs**
  - idea 1: embed graphs into vector space and use classical kernels



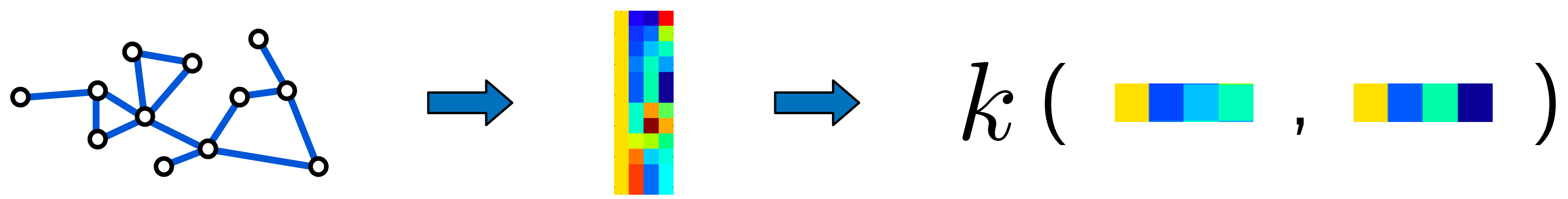
  - idea 2: use graph kernels [Borgwardt20]



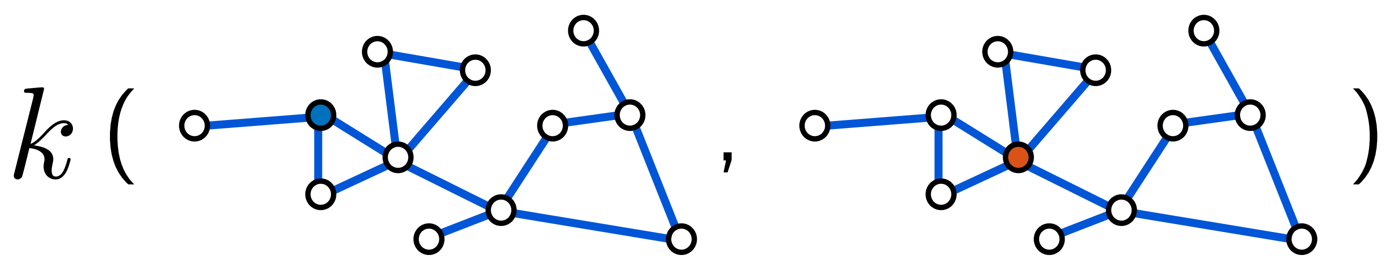
Borgwardt et al., "Graph kernels: State-of-the-art and future challenges," Now Publishers, 2020.

Kernels for graph-based functions

- Node-wise functions
  - key is comparison **between two nodes in the graph**
  - idea 1: embed nodes in the graph into vector space and use classical kernels



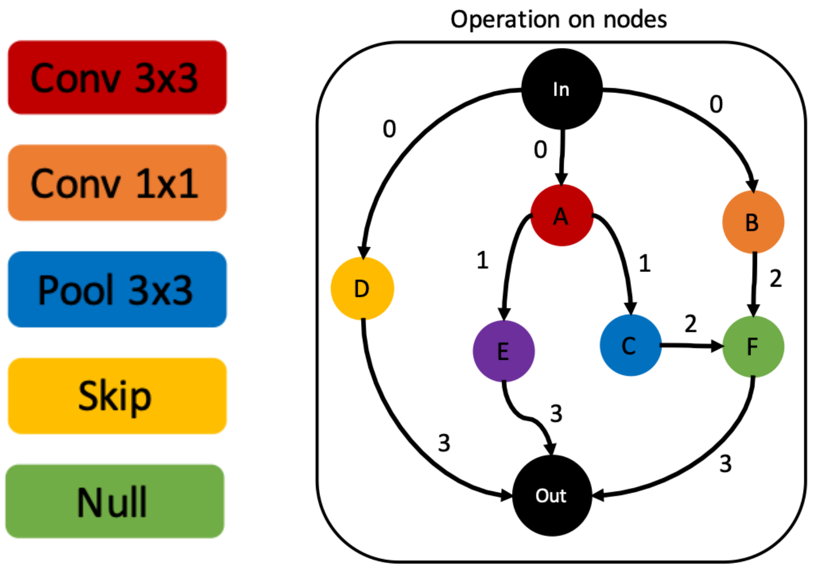
  - idea 2: use kernels defined on graphs [Smola03] (e.g., functions of graph Laplacian)



Smola and Kondor, "Kernels and regularization on graphs," COLT, 2003.

Case 1: Neural architecture search

- Objective: search for effective cell-based architecture for neural networks



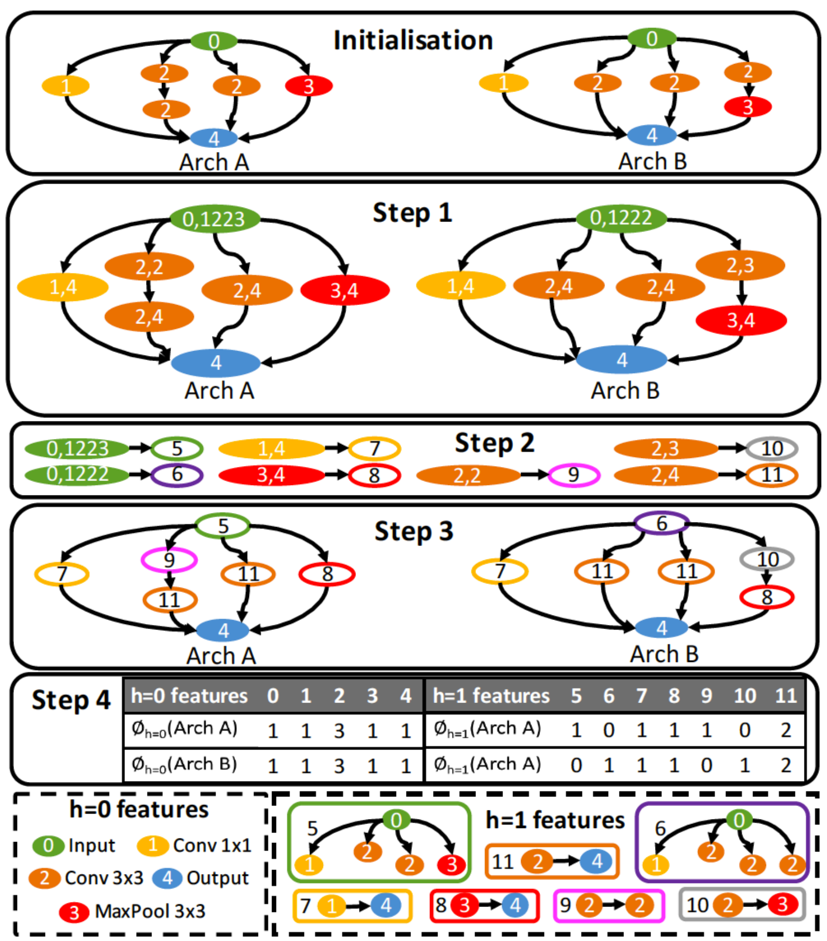
$$G^* = \arg\max_{G \in \mathcal{G}} y(G)$$

Algorithm 1 NAS-BOWL Algorithm.

Input: Maximum BO iterations T, BO batch size b, acquisition function $\alpha(\cdot)$, initial observed data on the target task $\mathcal{D}_0$, Optional: past-task query data $\mathcal{D}_{\text{past}}$ and surrogate $\mathcal{S}_{\text{past}}$
Output: The best architecture G^*_T
Initialise the GPWL surrogate $\mathcal{S}$ with $\mathcal{D}_0$
for $t = 1, \dots, T$ **do**
    Generate B candidate architectures $\mathcal{G}_t$
    $\{G_{t,i}\}_{i=1}^{b} = \arg\max_{G \in \mathcal{G}_t} \alpha_t(G|\mathcal{D}_{t-1})$
    Evaluate their validation accuracy $\{y_{t,i}\}_{i=1}^{b}$
    $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup (\{G_{t,i}\}_{i=1}^{B}, \{y_{t,i}\}_{i=1}^{b})$
    Update the surrogate $\mathcal{S}$ with $\mathcal{D}_t$
end for
Return the best architecture seen so far G^*_T

Ru et al., "Interpretable neural architecture search via Bayesian optimisation with Weisfeiler-Lehman kernels," ICLR, 2021.

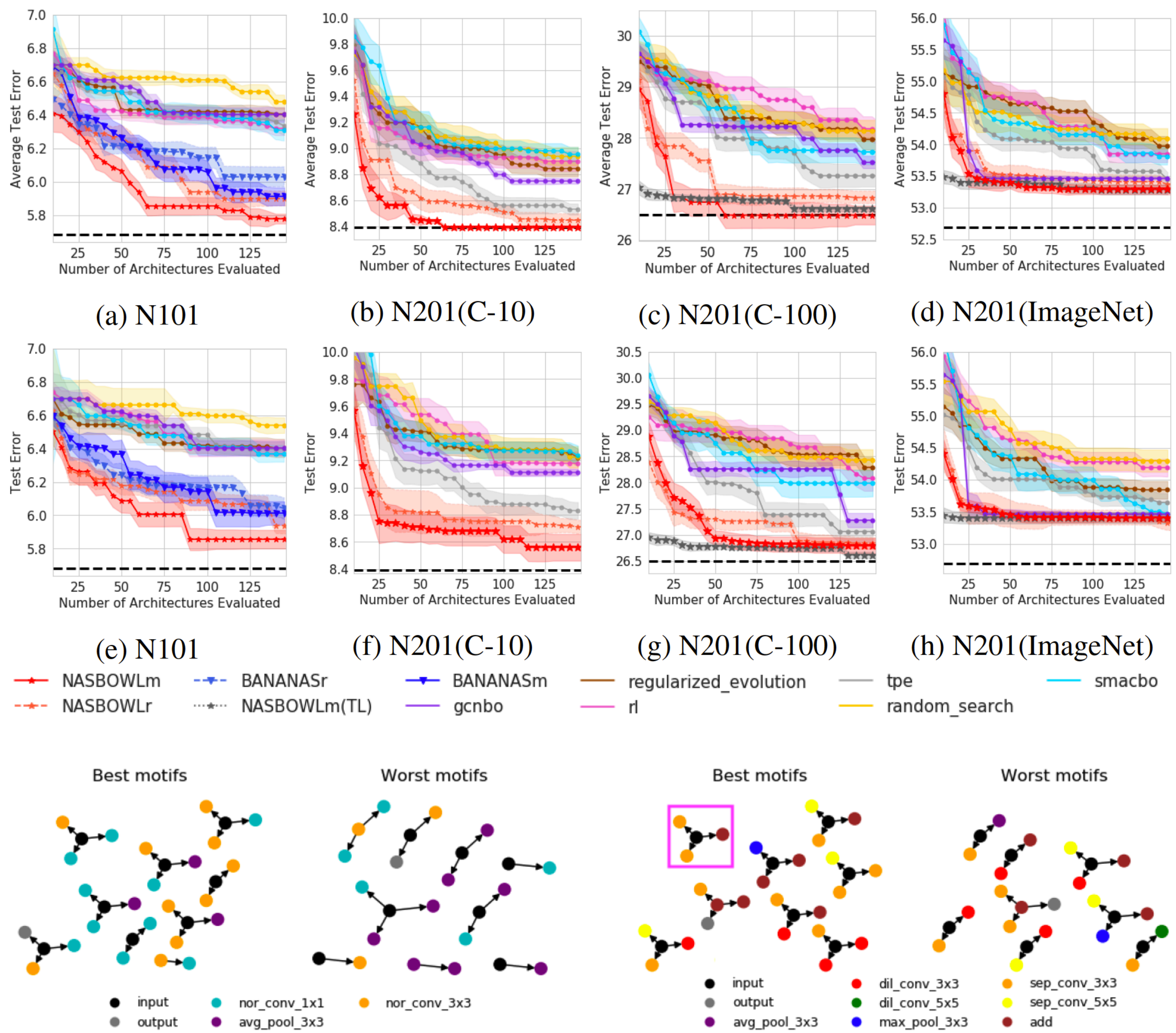
Case 1: Neural architecture search



Methodology

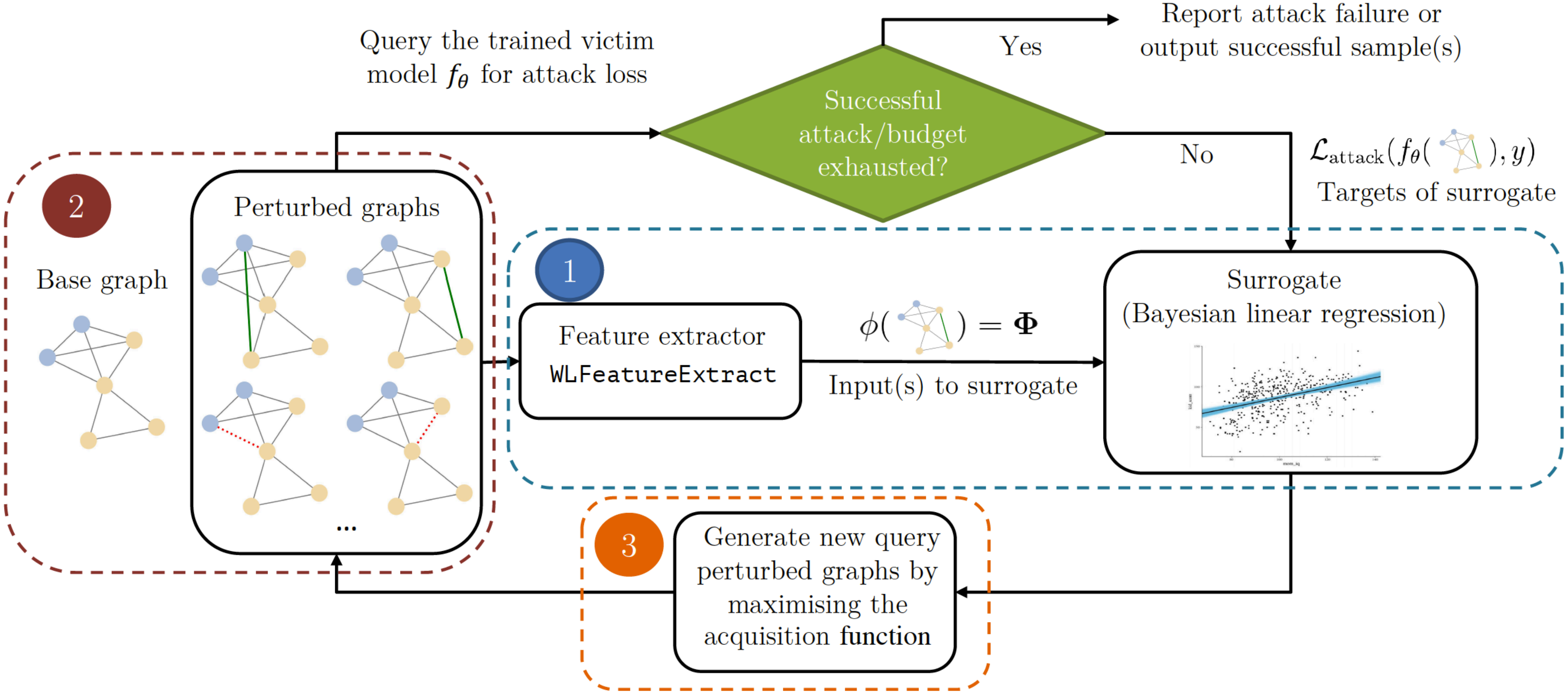
- surrogate model: Gaussian process via Weisfeiler-Lehman (WL) graph kernel measuring **similarity between graphs**
- candidate generation: mutation algorithm
- acquisition function: expected improvement (EI)

Case 1: Neural architecture search



Case 2: Adversarial attacks on graphs

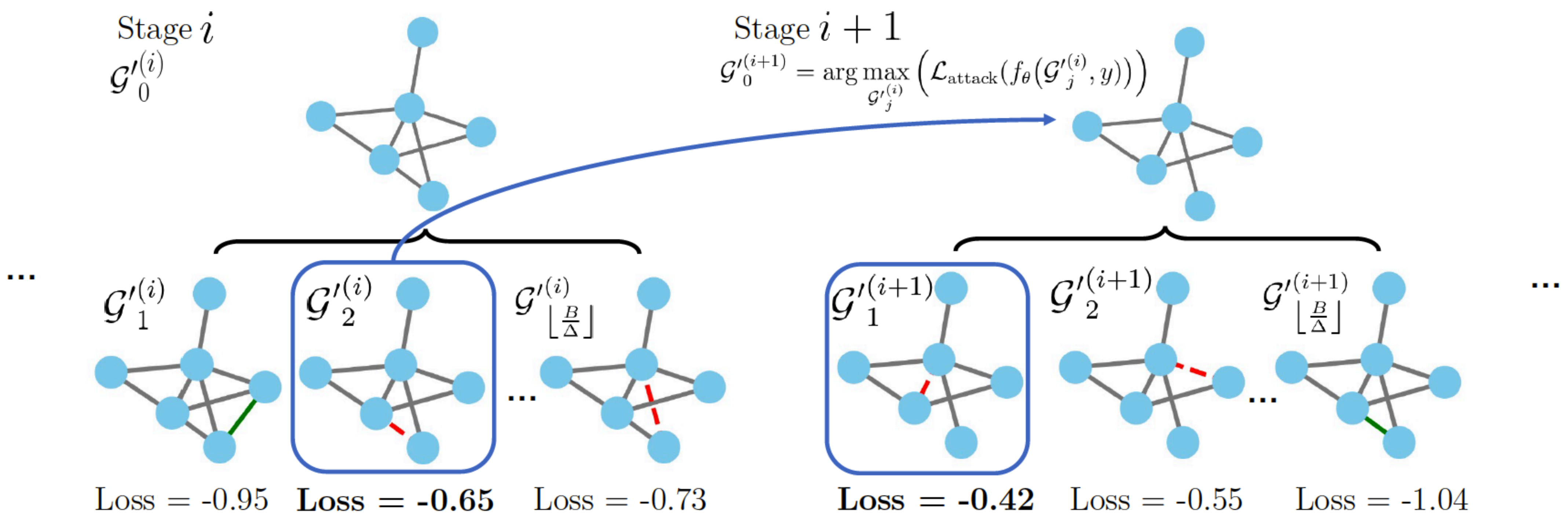
- Objective: find graph perturbation most harmful to trained graph classifier



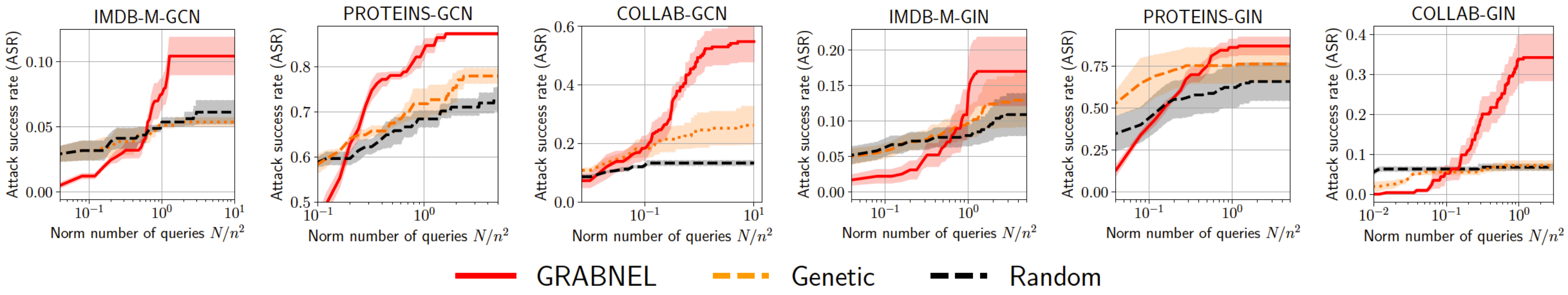
Wan et al., "Adversarial attacks on graph classifiers via Bayesian optimisation," NeurIPS, 2021.

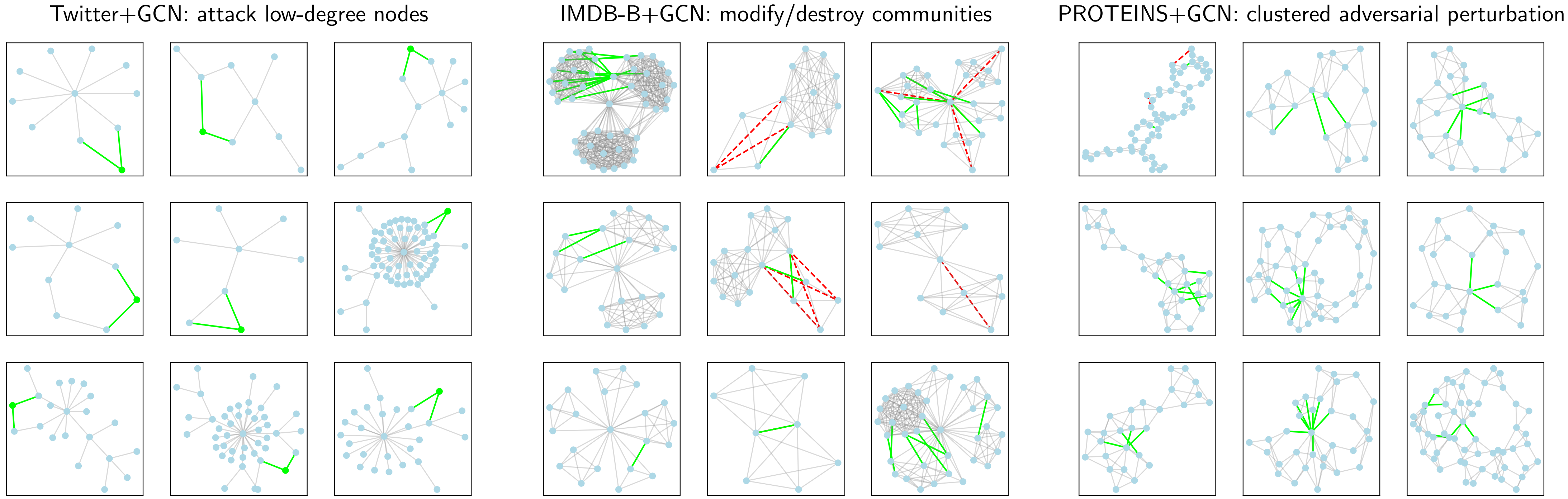
Case 2: Adversarial attacks on graphs

- Methodology
  - surrogate model: Bayesian linear regression with WL features
  - candidate generation: mutation by edge edit distance 1 from current evaluation
  - acquisition function: expected improvement (EI)



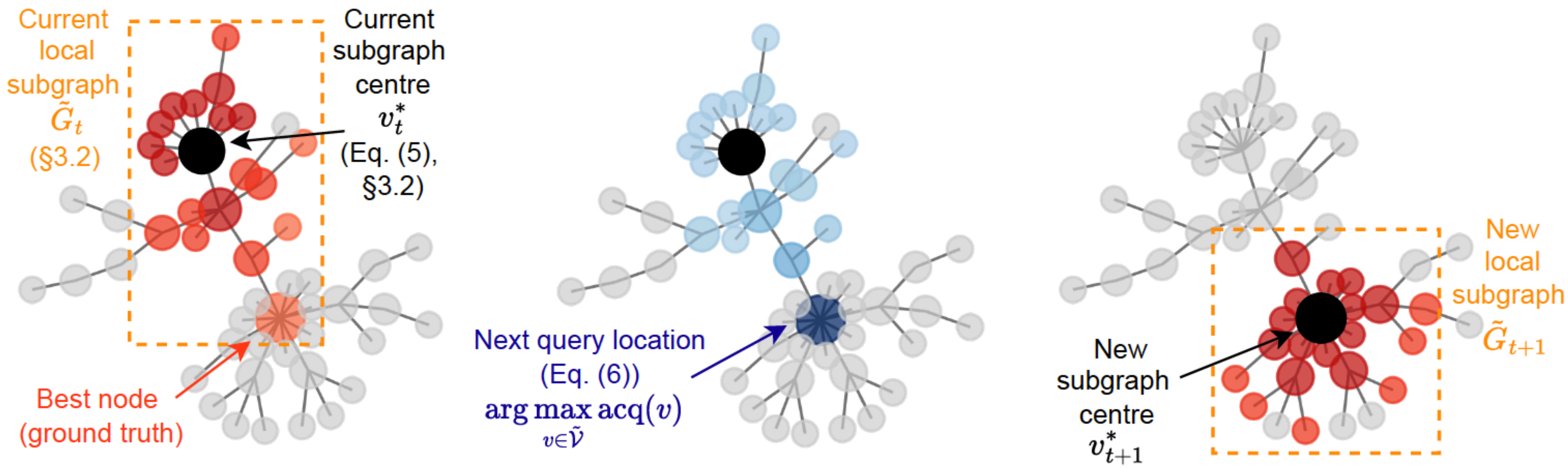
Case 2: Adversarial attacks on graphs





Case 3: Identifying most important user

- Objective: find the node with maximum value of a function on node set
- Methodology
  - progressively sample subgraph on-the-fly



Wan et al., "Bayesian optimisation of functions on graphs," NeurIPS, 2023.

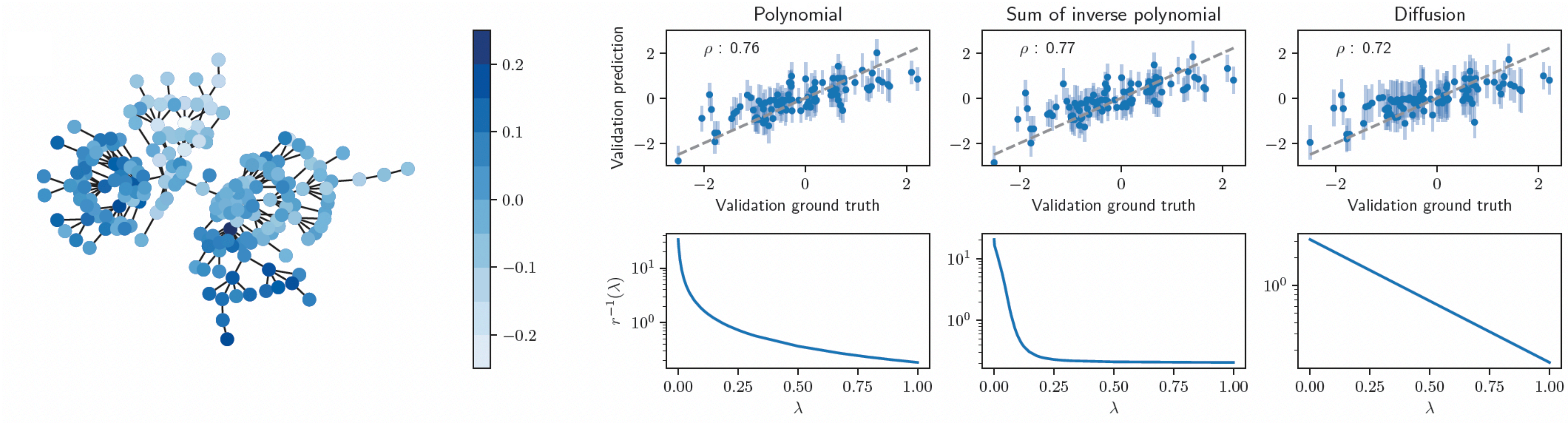
Case 3: Identifying most important user

- Objective: find the node with maximum value of a function on node set
- Methodology
  - progressively sample subgraph on-the-fly
  - surrogate model: Gaussian process via kernels measuring **similarity between nodes**

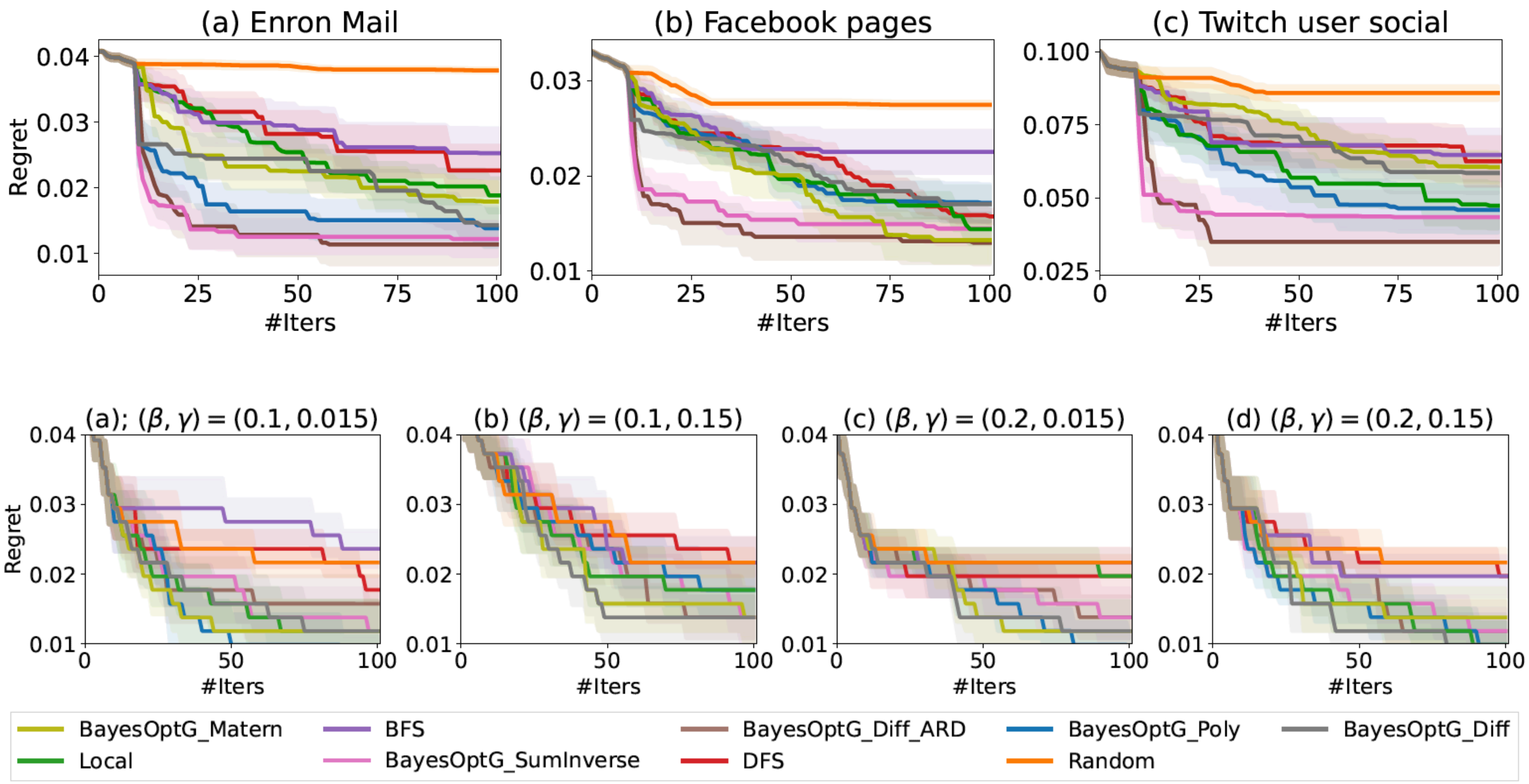
Kernel	Regularisation function $r(\lambda_i)$	Kernel function $K(\mathcal{V}, \mathcal{V})$
Diffusion† [37, 27]	$\exp(\beta_i \lambda_i)$	$\sum_{i=1}^{\tilde{n}} \exp(-\beta_i \lambda_i) \mathbf{u}_i \mathbf{u}_i^\top$
Polynomial*	$\sum_{\alpha=0}^{\eta-1} \beta_\alpha \lambda_i^\alpha + \epsilon$	$\sum_{i=1}^{\tilde{n}} \left(\sum_{\alpha=0}^{\eta-1} \beta_\alpha \lambda_i^\alpha + \epsilon \right)^{-1} \mathbf{u}_i \mathbf{u}_i^\top$
Sum-of-inverse polynomials*	$\left(\sum_{\alpha=0}^{\eta-1} \frac{1}{\beta_\alpha \lambda_i^\alpha + \epsilon} \right)^{-1}$	$\sum_{i=1}^{\tilde{n}} \left(\sum_{\alpha=0}^{\eta-1} \frac{1}{\beta_\alpha \lambda_i^\alpha + \epsilon} \right) \mathbf{u}_i \mathbf{u}_i^\top$
Matérn [4]	$\left(\beta\nu + \lambda_i \right)^{\nu}$	$\sum_{i=1}^{\tilde{n}} \left(\beta\nu + \lambda_i \right)^{-\nu} \mathbf{u}_i \mathbf{u}_i^\top$

  - acquisition function: expected improvement (EI)

Case 3: Identifying most important user



Case 3: Identifying most important user



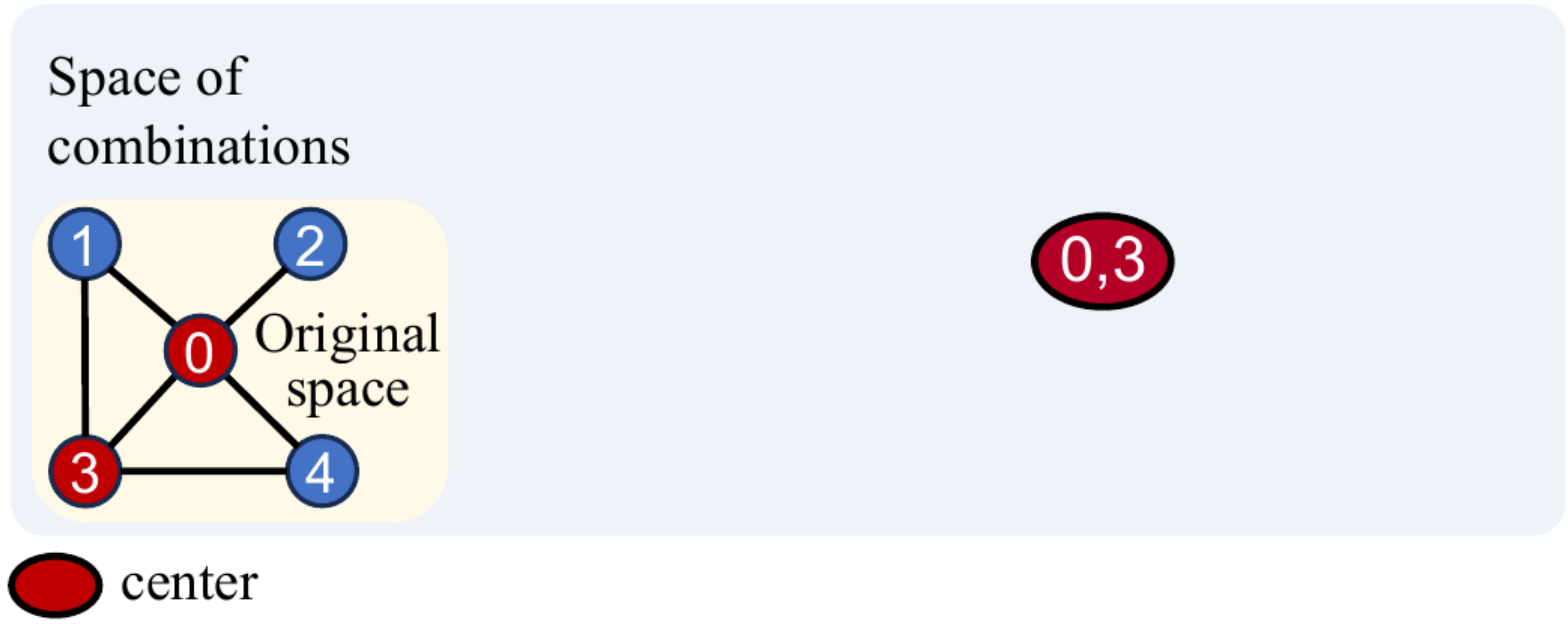
Case 4: Identifying subset of important nodes

- Objective: find the node subset with maximum value of a function
- Methodology
  - convert original graph into combo-graph where each node corresponds to a subset
  - follow methodology in case 3

Liang et al., "Bayesian optimization of functions over node subsets in graphs," NeurIPS, 2024.

Case 4: Identifying subset of important nodes

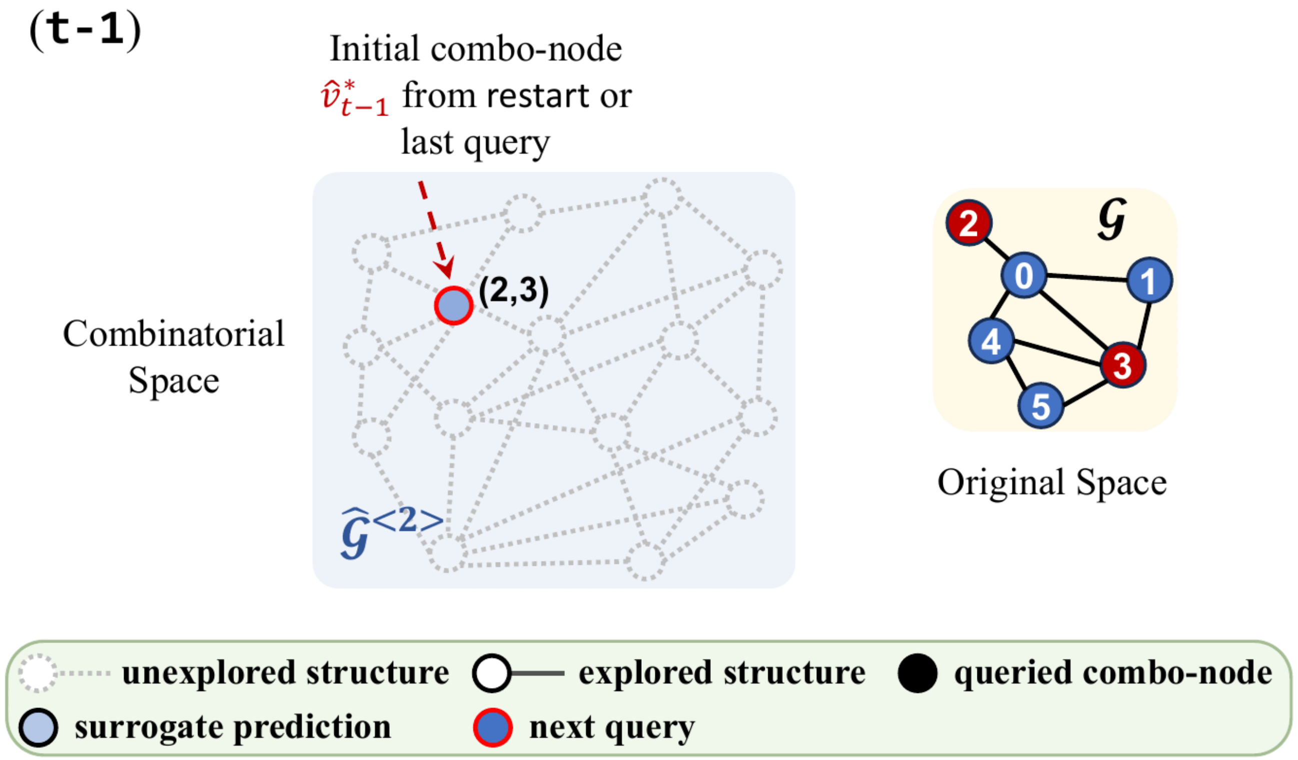
- Objective: find the node subset with maximum value of a function
- Methodology
  - convert original graph into combo-graph where each node corresponds to a subset
  - follow methodology in case 3



Liang et al., "Bayesian optimization of functions over node subsets in graphs," NeurIPS, 2024.

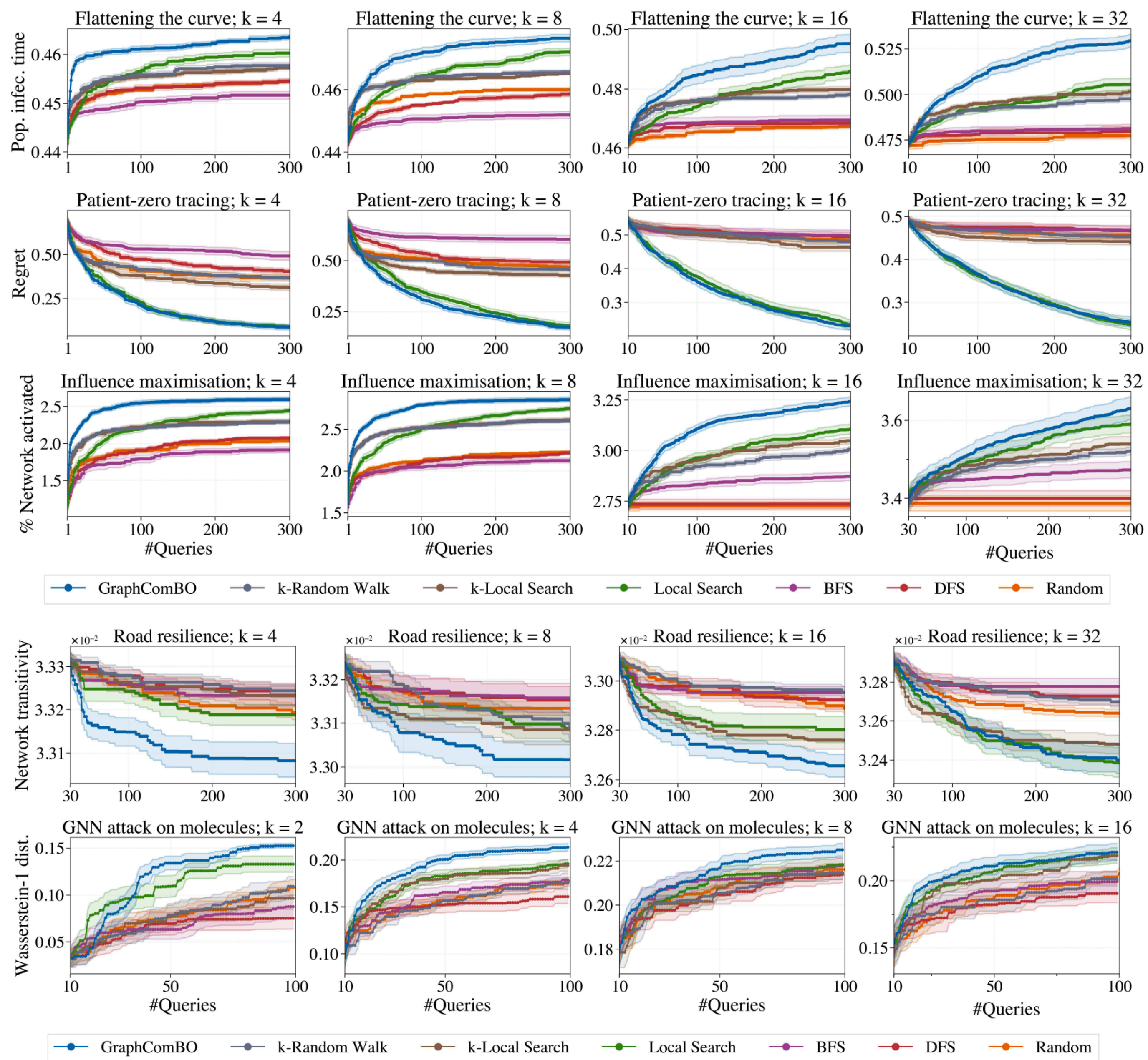
Case 4: Identifying subset of important nodes

- Objective: find the node subset with maximum value of a function
- Methodology
  - convert original graph into combo-graph where each node corresponds to a subset
  - follow methodology in case 3



Liang et al., "Bayesian optimization of functions over node subsets in graphs," NeurIPS, 2024.

Case 4: Identifying subset of important nodes



Discussion

- Combination of Bayesian and graph modelling
  - Bayesian modelling provides probabilistic perspective
  - graph modelling provides geometric/topological perspective
- New and interesting problems
  - Optimisation of graph-based functions
  - Gaussian processes/uncertainty estimation on graphs
  - Graph structure optimisation for downstream tasks
- Open challenges
  - computational complexity and scalability
  - theoretical analysis

References

- B. Ru, X Wan, X. Dong and M. A. Osborne, "Interpretable neural architecture search via Bayesian optimisation with Weisfeiler-Lehman kernels," *International Conference on Learning Representations (ICLR)*, 2021.
- X. Wan, H. Kenlay, B. Ru, A. Blaas, M. A. Osborne and X. Dong, "Adversarial attacks on graph classifiers via Bayesian optimisation," *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- X. Wan, P. Osselin, H. Kenlay, B. Ru, M. A. Osborne and X. Dong, "Bayesian optimisation of functions on graphs," *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- H. Liang, X. Wan and X. Dong, "Bayesian optimization of functions over node subsets in graphs," *Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- F. L. Opolka, Y.-C. Zhi, P. Liò and X. Dong, "Adaptive Gaussian processes on graphs via spectral graph wavelets," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.
- Y.-C. Zhi, Y. C. Ng and X. Dong, "Gaussian processes on graphs via spectral kernel learning," *IEEE Transactions on Signal and Information Processing over Networks (TSIPN)*, 2023.
- F. L. Opolka, Y.-C. Zhi, P. Liò and X. Dong, "Graph classification Gaussian processes via spectral features," *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2023.
- H. Sáez de Ocáriz Borde, Á. Arroyo, I. Morales, I. Posner and X. Dong, "Neural latent geometry search: Product manifold inference via Gromov-Hausdorff-informed Bayesian optimization," *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.