Halting in Random Walk Kernels

Mahito Sugiyama (Osaka University, JST PRESTO) Karsten Borgwardt (ETH Zürich)

NIPS2015, December 7–12, 2015

Code: `https://www.bsse.ethz.ch/mlcb/research/machine-learning/graph-kernels.html`

D-BSSE Department of Biosystems Science and Engineering

MLCB

Our Messages

1. As a baseline for graph kernels, a *fixed-length-k random walk kernel* is better than a *geometric random walk kernel*
2. ***Simple baseline kernels on label histograms*** should be employed

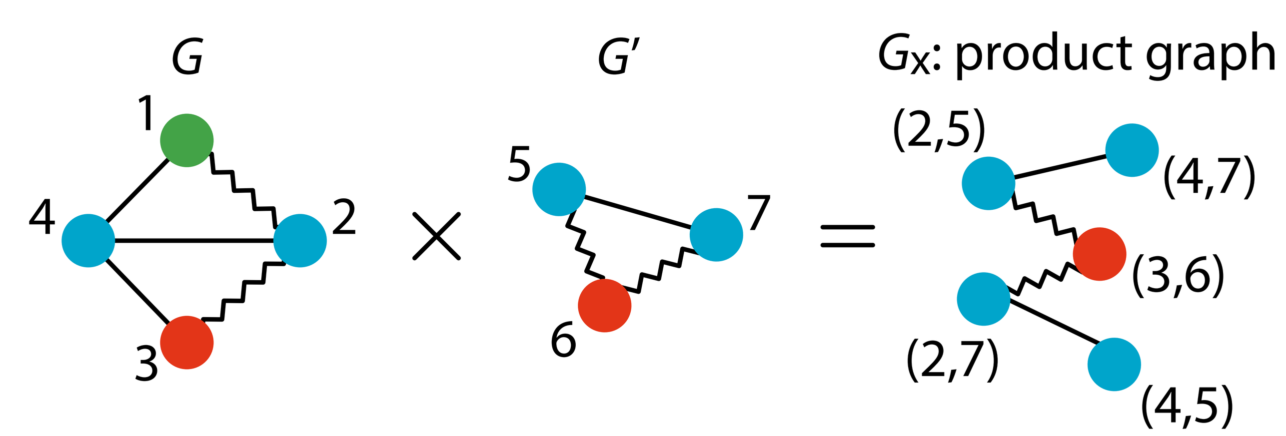
Random Walk Kernels

- Measure the similarity between graphs by counting matching walks
- The direct product $G_\times = (V_\times, E_\times, \varphi_\times)$ of two graphs G and G':

$$V_\times = \{ (v, v') \in V \times V' \mid \varphi(v) = \varphi'(v') \},$$

$$E_\times = \left\{ ((u, u'), (v, v')) \in V_\times \times V_\times \;\middle|\; \begin{array}{l} (u, v) \in E, \\ (u', v') \in E', \\ \varphi(u, v) = \varphi'(u', v') \end{array} \right\}$$

  - All labels are inherited



- The k-step (fixed-length-k) random walk kernel between G and G':

$$K_\times^k(G, G') = \sum_{i,j=1}^{|V_\times|} \left[\lambda_0 A_\times^0 + \lambda_1 A_\times^1 + \lambda_2 A_\times^2 + \cdots + \lambda_k A_\times^k \right]_{ij} \quad (\lambda_l > 0)$$

  - $A_\times$: The adjacency matrix of the product graph

- $K_\times^\infty$ can be directly computed if $\lambda_\ell = \lambda^\ell$ for each $\ell \in \{0, \dots, k\}$ (geometric series), resulting in the geometric random walk kernel:

$$K_{\mathrm{GR}}(G, G') = \sum_{i,j=1}^{|V_\times|} \left[\lambda^0 A_\times^0 + \lambda^1 A_\times^1 + \lambda^2 A_\times^2 + \lambda^3 A_\times^3 + \cdots \right]_{ij} = \sum_{i,j=1}^{|V_\times|} \left[\sum_{\ell=0}^{\infty} \lambda^\ell A_\times^\ell \right]_{ij}$$

$$= \sum_{i,j=1}^{|V_\times|} \left[(\mathbf{I} - \lambda A_\times)^{-1} \right]_{ij}$$

  - Well-defined only if $\lambda < 1/\mu_{\times,\max}$ ($\mu_{\times,\max}$ is the max. eigenvalue of $A_\times$)
  - $\delta_\times$ (min. degree) $\le \bar{d}_\times$ (average degree) $\le \mu_{\times,\max} \le \Delta_\times$ (max. degree)

Main Theorem

- Since λ is relatively small, halting of random walks occurs:

$$K_{\mathrm{GR}}(G, G') = \sum_{i,j=1}^{|V_\times|} [\, \underbrace{\lambda^0 A_\times^0 + \lambda^1 A_\times^1}_{K_\times^1(G,G')} + \underbrace{\lambda^2 A_\times^2 + \lambda^3 A_\times^3 + \cdots}_{\to 0} \,]_{ij}$$

- **Theorem**: For a pair of graphs G and G',

$$K_\times^1(G, G') \le K_{\mathrm{GR}}(G, G') \le K_\times^1(G, G') + \varepsilon, \qquad \varepsilon = |V_\times| \frac{(\lambda \Delta_\times)^2}{1 - \lambda \Delta_\times}$$

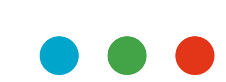
  - $\varepsilon \to 0$ (monotonic) as $\lambda \to 0$
  - $\lambda_0 = 1$ and $\lambda_1 = \lambda$ in the random walk kernel
  - Normalized version:

$$1 \le \frac{K_{\mathrm{GR}}(G, G')}{K_\times^1(G, G')} \le 1 + \varepsilon', \qquad \varepsilon' = \frac{(\lambda \Delta_\times)^2}{(1 - \lambda \Delta_\times)(1 + \lambda \bar{d}_\times)}$$

Relationships to Linear Kernels

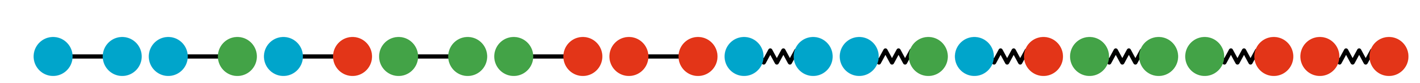
- The lower bound $K_\times^1(G, G')$ is just a linear kernel on label histograms:

$$K_\times^1(G, G') = \underbrace{K_{\mathrm{VH}}(G, G')}_{\text{Vertex labels}} + \lambda \underbrace{K_{\mathrm{VEH}}(G, G')}_{\text{Vertex + edge labels}}$$



G	2	1	1
G'	2	0	1

$K_{\mathrm{VH}}(G, G') = 5$



G	1	1	1	0	0	0	0	1	1	0	0	0
G'	1	0	0	0	0	0	0	0	2	0	0	0

$K_{\mathrm{VEH}}(G, G') = 3$

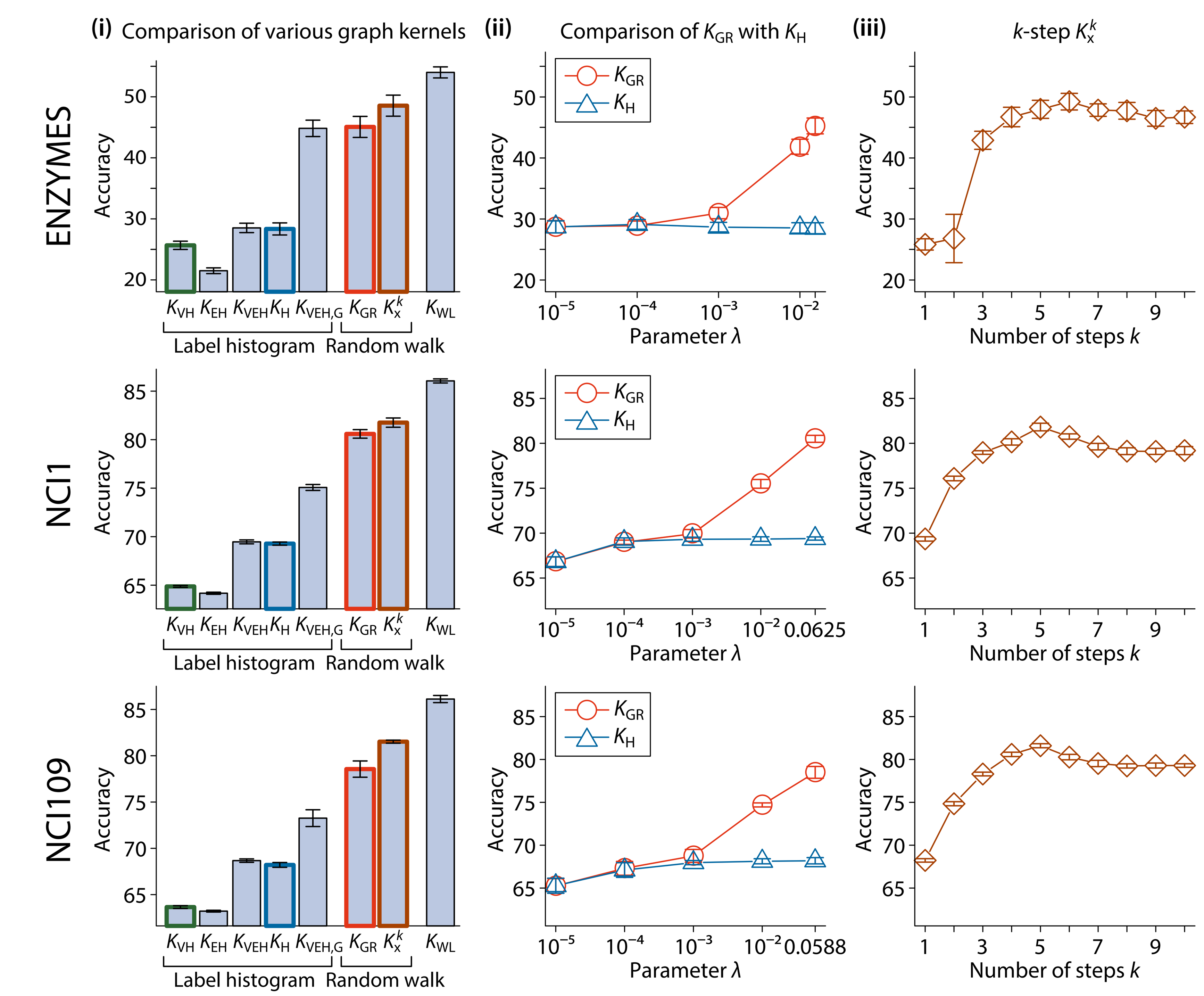
Consequence

Geometric random walk kernels may degenerate to simple kernels between node and edge label histograms

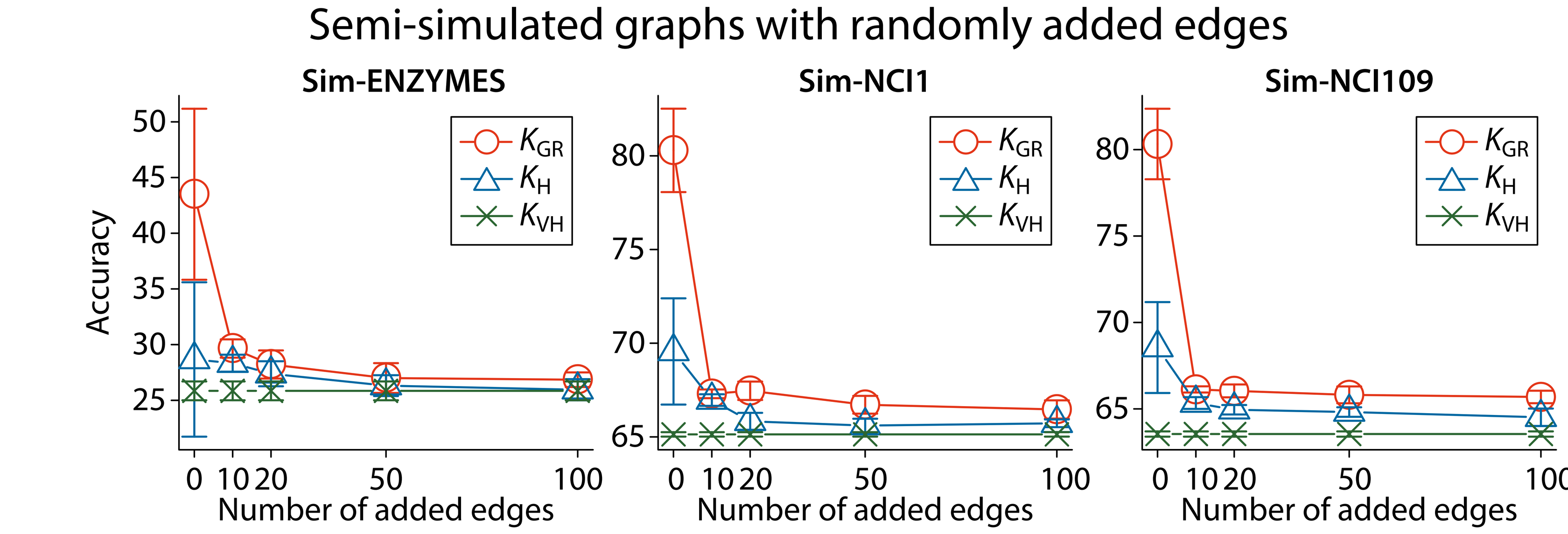
Datasets

Dataset	Size	#cls.	avg.$\lvert V\rvert$	avg.$\lvert E\rvert$	max$\lvert V\rvert$	max$\lvert E\rvert$	$\lvert\Sigma_V\rvert$	$\lvert\Sigma_E\rvert$	max$\Delta_\times$
ENZYMES	600	6	32.63	62.14	126	149	3	1	65
NCI1	4110	2	29.87	32.3	111	119	37	3	16
NCI109	4127	2	29.68	32.13	111	119	38	3	17

Experimental Results



Semi-simulated graphs with randomly added edges



References

- Borgwardt, K. M.: ***Graph Kernels***. PhD thesis, 2007.
- Gärtner, T., Flach, P., and Wrobel, S.: **On graph kernels: Hardness results and efficient alternatives.** In *Learning Theory and Kernel Machines* (*LNCS* 2777), 129–143, 2003.
- Kashima, H., Tsuda, K., and Inokuchi, A.: **Marginalized kernels between labeled graphs.** *ICML*, 321–328, 2003.
- Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M.: **Weisfeiler-Lehman graph kernels.** *JMLR*, 12:2359–2561, 2011.