Is IRC-safe information all you need for jet classification?

Dimitrios Athanasakos

arXiv: 2305.08979 w/ Andrew J. Larkoski, **James Mulligan**, Mateusz Płoskoń, **Felix Ringer**

CFNS, AI at the EIC

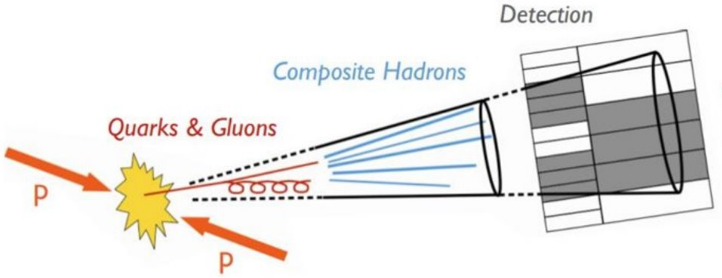
Stony Brook University

Outline

Table of Contents

Jet Classification and IRC-Safety

One of the biggest challenges of collider phenomenology is Jet Classification



Jet Classification and IRC-Safety

There are many ways to represent a jet

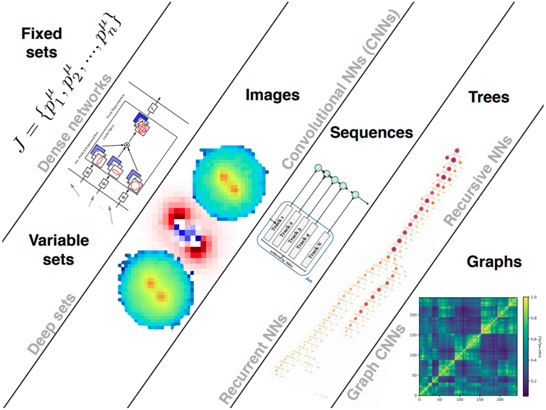
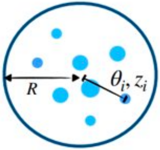


Figure: Taken from Larkoski et al 1709.04464

Jet Classification and IRC-Safety

We are free to construct any observable from the jet's constituents

e.g. $$\lambda_\alpha^\kappa = \sum_{i \in \mathrm{jet}} z_i^\kappa \theta_i^\alpha$$

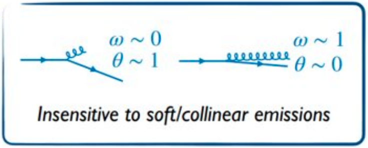


$$\theta_i = \frac{\sqrt{\Delta y^2 + \Delta\varphi^2}}{R}$$

$$z_i = \frac{p_{\mathrm{T},i}}{p_{\mathrm{T,jet}}}$$

However, usually only those combinations that obey **infrared-collinear (IRC) safety** are calculable in perturbative QCD

e.g. $$\lambda_{\alpha>0}^{\kappa=1} = \sum_{i \in \mathrm{jet}} z_i \theta_i^\alpha$$



Jet Classification and IRC-Safety

Architectures that use IRC unsafe information (PFN, ParticleNet etc) perform better than IRC safe classifiers (EFN, EFP, Nsub)

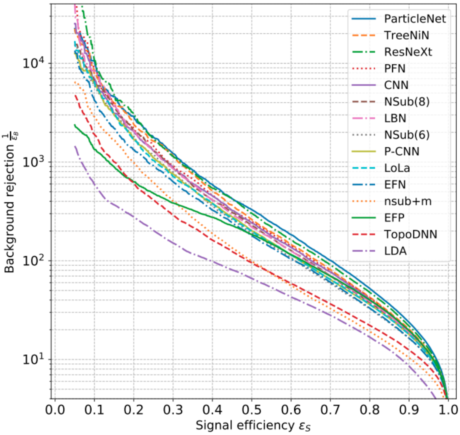


Figure: Taken from Kasieczka et al 1902.09914

Deep Sets

Permutation-invariant neural networks based on deep sets

Unordered, variable-length sets of particles as input

Komiske, Metodiev, Thaler JHEP 01 (2019) 121

Zaheer et al. 1703.06114
Wagstaff et al. 1901.09006
Bloem-Reddy, Teh JMLR 21 90 (2020)

Particle Flow Network (PFN)

$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} \Phi\left(p_i\right)\right)$$

Classifier — DNNs — latent space $d = 256$

Includes IRC-unsafe information

Energy Flow Network (EFN)

$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} z_i \Phi\left(\hat{p}_i\right)\right)$$

Classifier — DNNs

Includes only IRC-safe information

Deep Sets

PFN performs amazingly well and almost matches the state of the art performance

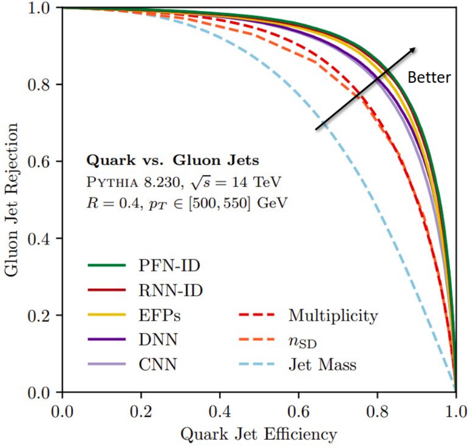


Figure: Taken from Thaler et al 1810.05165

Jet Classification and IR-Safety

Interpretability

PFN is IRC unsafe, sensitive to non perturbative physics and it has 3N variables where N is the number of hadrons

→

Increase interpretability by connecting it to Sudakov/IRC safe observables and by cutting down the input's size

Table of Contents

JFN

Jet Flow Network (JFN)

Instead of hadrons, we work with subjets

↓

For a non-zero subjet radius r the classifier is Sudakov safe

↓

As we lower r, we smoothly transition to the non-perturbative regime

↓

Properties of inclusive and exclusive subjets are known to NLL *(L. Dai et al 1606.07411, F.Ringer et al 1705.05375, 2103.16573, A. Larkoski et al 2205.01117)*

	PFN	JFN	EFN
Input	particle 3-momenta	subjet 3-momenta	particle 3-momenta
Classifier	IRC unsafe	Sudakov safe	IRC safe

JFN

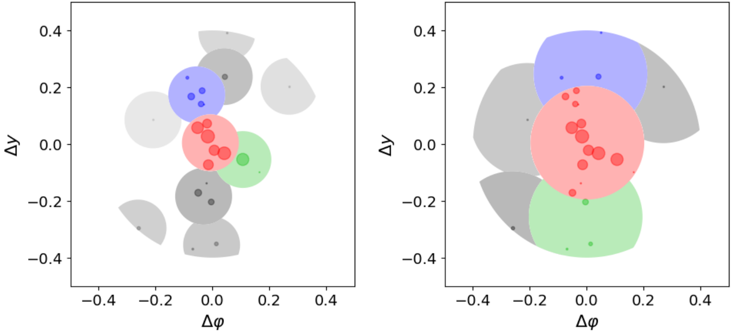


Figure: A QCD jet with $p_T = 100$ GeV, $R = 0.4$ reclustered into subjets for subjet radii $r = 0.1$ (left), $r = 0.2$ (right). The radii of the particles represent their p_T. Leading subjet: red. Second leading subjet: green. Third leading subjet: blue.

JFN

As we lower r JFN converges to PFN

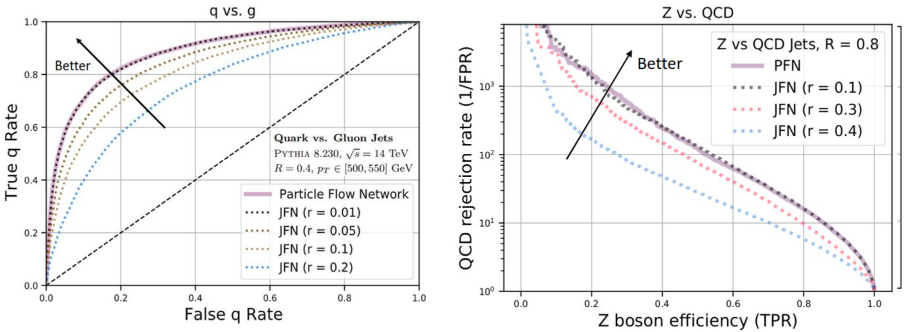
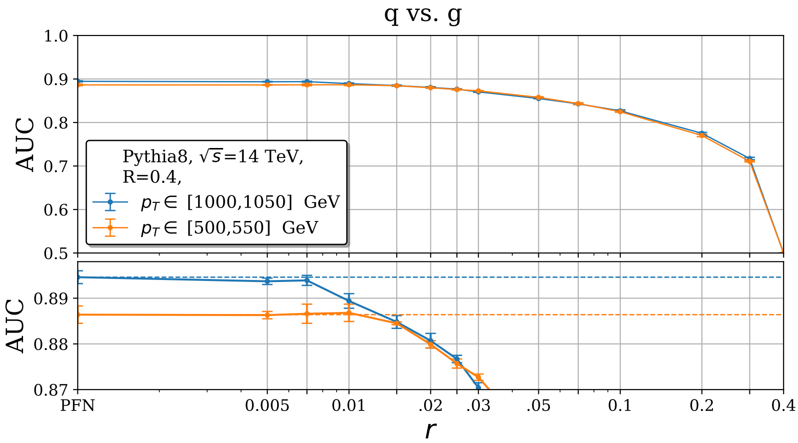


Table of Contents

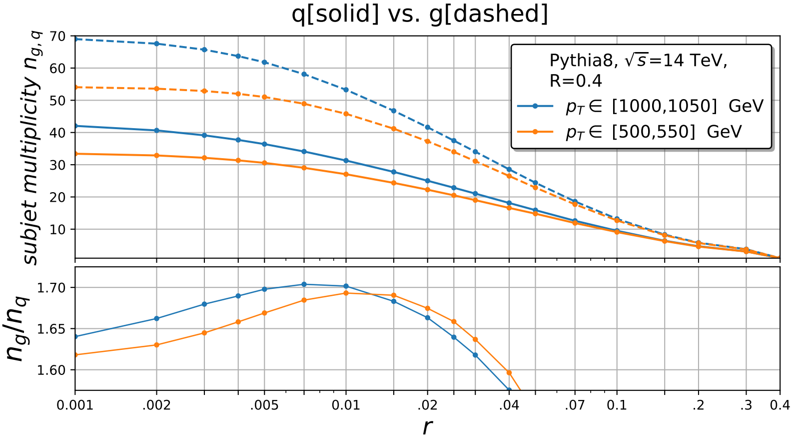
Physical Scales

At what r do we expect a decrease in performance ?



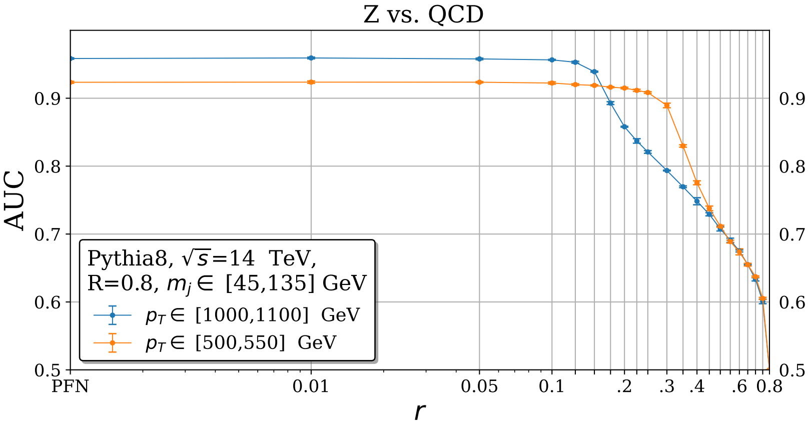
Physical Scales

The critical scale is not the hadronization scale $\sim 0.5\ GeV$.
It corresponds to $p_T \cdot r \sim 5\ GeV$



Physical Scales

Z vs QCD Discrimination



Physical Scales

What happens at this scale ?

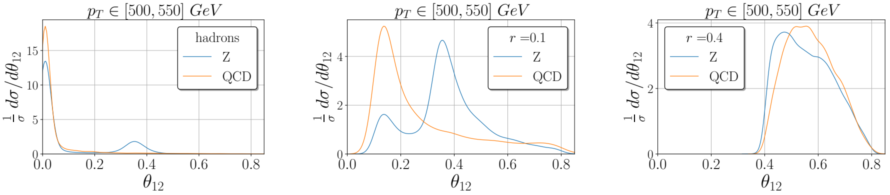


Figure: The distribution of the angle θ in the (η, ϕ) plane between the two leading hadrons (left plot) and the two leading subjets (center and right plot).

Physical Scales

We do not lose classification power as long as we resolve the two leading subjets that originate from the Z splitting: $\theta_Z \approx \frac{2M_Z}{p_T}$

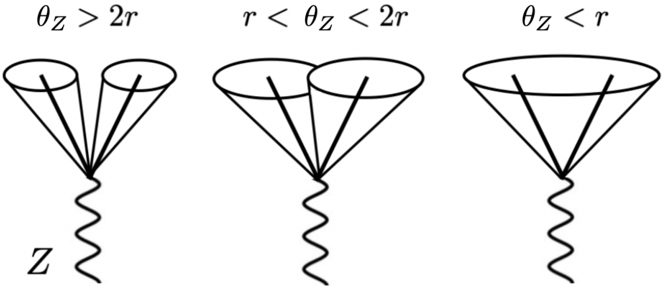


Table of Contents

Conclusion

In conclusion:

- Jet Flow Network is a Sudakov Safe classifier

Conclusion

In conclusion:

- Jet Flow Network is a Sudakov Safe classifier
- The first classifier based on IRC safe information that matches the performance of an IRC-unsafe one with the same expressive power

Conclusion

In conclusion:

- Jet Flow Network is a Sudakov Safe classifier
- The first classifier based on IRC safe information that matches the performance of an IRC-unsafe one with the same expressive power
- Increased interpretability (fewer variables, connections to pQCD)

Conclusion

In conclusion:

- Jet Flow Network is a Sudakov Safe classifier
- The first classifier based on IRC safe information that matches the performance of an IRC-unsafe one with the same expressive power
- Increased interpretability (fewer variables, connections to pQCD)
- Decreased dependency to Monte Carlo Hadronization Models (improved robustness?)

Table of Contents

Minimal Information and Graph Theory

- JFN is an important step towards increasing the interpretability of a Jet Classifier without sacrificing on the performance.

Minimal Information and Graph Theory

- JFN is an important step towards increasing the interpretability of a Jet Classifier without sacrificing on the performance.
- That was achieved by using (p_T, η, ϕ) of subjets as input to a Deep Sets classifier.

Minimal Information and Graph Theory

- JFN is an important step towards increasing the interpretability of a Jet Classifier without sacrificing on the performance.
- That was achieved by using (p_T, η, ϕ) of subjets as input to a Deep Sets classifier.
- Can we do better?

Minimal Information and Graph Theory

- The η and ϕ of individual hadrons/subjets are not QCD observables, only the relative angles ΔR between them are.

Minimal Information and Graph Theory

- The $\eta \ and \ \phi$ of individual hadrons/subjets are not QCD observables, only the relative angles ΔR between them are.
- For a configuration of N particles, the phase space is $3N - 3$ dimensional. We can split the phase space in N transverse momenta and $2N - 3$ relative angles. (*1704.08249, 2111.14589, 2008.06508*)

Minimal Information and Graph Theory

- The $\eta \; and \; \phi$ of individual hadrons/subjets are not QCD observables, only the relative angles ΔR between them are.
- For a configuration of N particles, the phase space is $3N-3$ dimensional. We can split the phase space in N transverse momenta and $2N-3$ relative angles. (*1704.08249, 2111.14589, 2008.06508*)
- Let's create a jet classifier where the input is a graph $G=(V,E)$ with $|V|=N, |E|=2N-3$ and $V, E \in \mathbb{R}$. In principle this graph contains all the necessary information to construct the phase space and the input features are all QCD observables.

Minimal Information and Graph Theory

- The η and ϕ of individual hadrons/subjets are not QCD observables, only the relative angles ΔR between them are.
- For a configuration of N particles, the phase space is $3N-3$ dimensional. We can split the phase space in N transverse momenta and $2N-3$ relative angles. (*1704.08249, 2111.14589, 2008.06508*)
- Let's create a jet classifier where the input is a graph $G=(V,E)$ with $|V|=N, |E|=2N-3$ and $V, E \in \mathbb{R}$. In principle this graph contains all the necessary information to construct the phase space and the input features are all QCD observables.
- How easy is it to reconstruct the phase space ?

Minimal Information and Graph Theory

Lets consider a similar (and harder) problem

- Given a graph $G = (V, E)$ with $|V| = N$ and a set of non-negative edge-weights $\{w_{ij} : (i, j) \in E\}$. The vertices don't carry any information.

Minimal Information and Graph Theory

Lets consider a similar (and harder) problem

- Given a graph $G = (V, E)$ with $|V| = N$ and a set of non-negative edge-weights $\{w_{ij} : (i, j) \in E\}$. The vertices don't carry any information.
- Can we find a realization of G in $\mathbb{R}^2$, i.e. can we calculate the distances between every pair of vertices ?

Minimal Information and Graph Theory

Lets consider a similar (and harder) problem

- Given a graph $G = (V, E)$ with $|V| = N$ and a set of non-negative edge-weights $\{w_{ij} : (i, j) \in E\}$. The vertices don't carry any information.
- Can we find a realization of G in $\mathbb{R}^2$, i.e. can we calculate the distances between every pair of vertices ?

NO

There are two issues

- In order to have a unique realization we need $|E| \geq 2N - 2$ (Global Rigidity in $\mathbf{R}^2$)

There are two issues

- In order to have a unique realization we need $|E| \geq 2N - 2$ (Global Rigidity in $\mathbf{R}^2$)
- Even then, this is an NP-hard problem (*Graph Realization Problem*)

There are two issues

- In order to have a unique realization we need $|E| \geq 2N - 2$ (Global Rigidity in $\mathbf{R}^2$)
- Even then, this is an NP-hard problem (*Graph Realization Problem*)
- The best we can do with $|E| = 2N - 3$ is Laman Graphs, which are graphs that have only finitely many realizations.

There are two issues

- In order to have a unique realization we need $|E| \geq 2N - 2$ (Global Rigidity in $\mathbf{R}^2$)
- Even then, this is an NP-hard problem (*Graph Realization Problem*)
- The best we can do with $|E| = 2N - 3$ is Laman Graphs, which are graphs that have only finitely many realizations.
- Formally, a Laman graph is a graph on N vertices such that, for all k, every k-vertex subgraph has at most $2k - 3$ edges, and such that the whole graph has exactly $2N - 3$ edges.

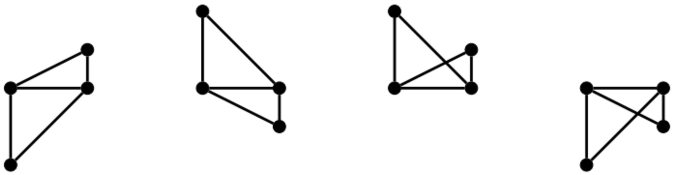
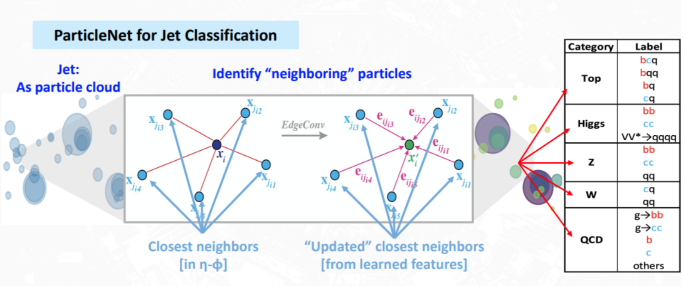


Figure: Realizations of a Laman graph with $|V| = 4$ up to rotations and translations.

Preliminary Results with ParticleNet



Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$\log p_T$	logarithm of the particle's p_T
$\log E$	logarithm of the particle's energy
$\log \frac{p_T}{p_T(\text{jet})}$	logarithm of the particle's p_T relative to the jet p_T
$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy
ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)

Figure: Input Variables to Particle Net

Preliminary Results with ParticleNet

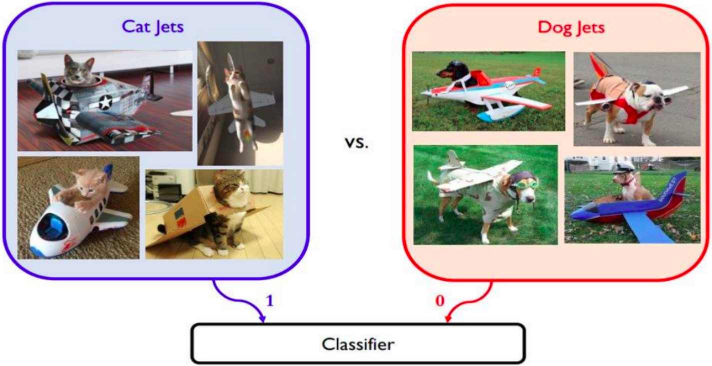
By forcing the ParticleNet to use Laman Graphs we can almost match the full performance, even though the input is $3N - 3$ dimensional compared to $21N$ dimensional.

Model	**AUC q vs g**
ParticleNet Lite $k = (7, 7)$	*0.8989*
Laman Graphs $k = (2, 7)$	*0.8983*
ParticleNet Lite $k = (2, 7)$	*0.8976*
PFN	*0.8910*

Table:

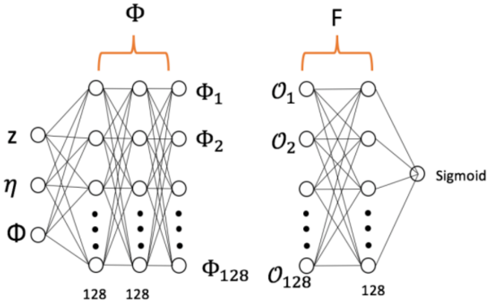
AUC of different classifiers on the q vs. g discrimination task. $k = (k_1, k_2)$ *refers to the number of nearest neighbors for the two convolutional layers of the architecture.*

Thank you!



Back up

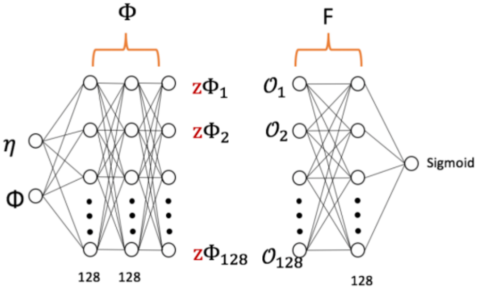
PFN: $F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$



Where $\mathcal{O}_a = \sum_i \Phi(z_i, \eta_i, \phi_i)$

Back up

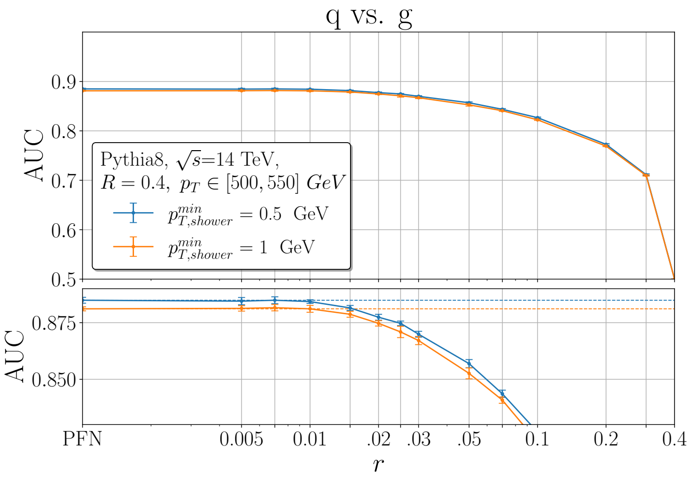
EFN: $F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right)$



Where $\mathcal{O}_a = \sum_i z_i \Phi(\eta_i, \phi_i)$

Back up

The location of the critical scale is independent of the $p_{T,shower}^{min}$ of Pythia



Back up

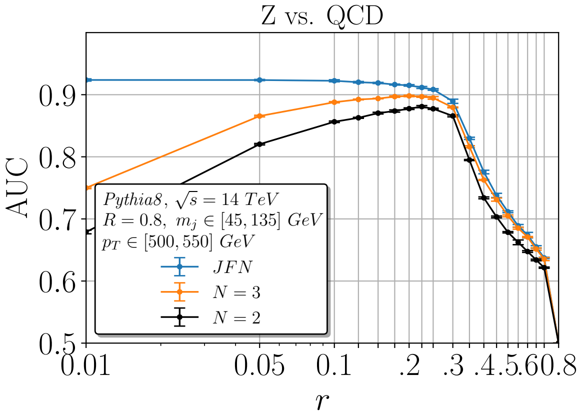


Figure: AUC of the JFNs for QCD vs. Z jets trained on the full information (inclusive subjets) compared to deep sets trained only on the two or three leading subjets.

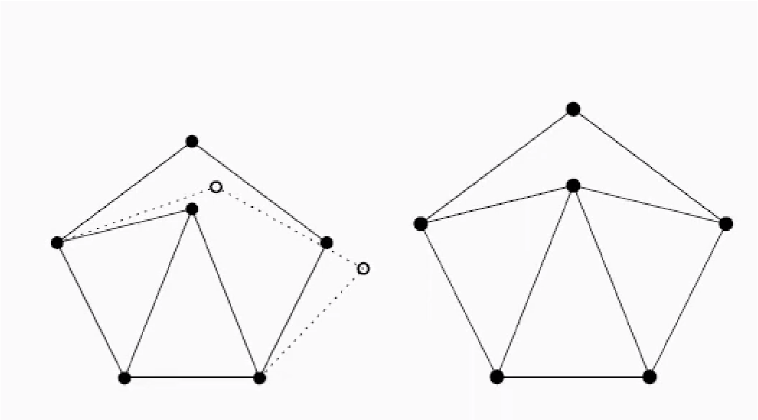


Figure: The left graph has infinitely many realizations. The right one is a Laman graph with only two realizations.