ML4Jets 2022, 1-4 Nov

Infrared and collinear safe graph neural networks

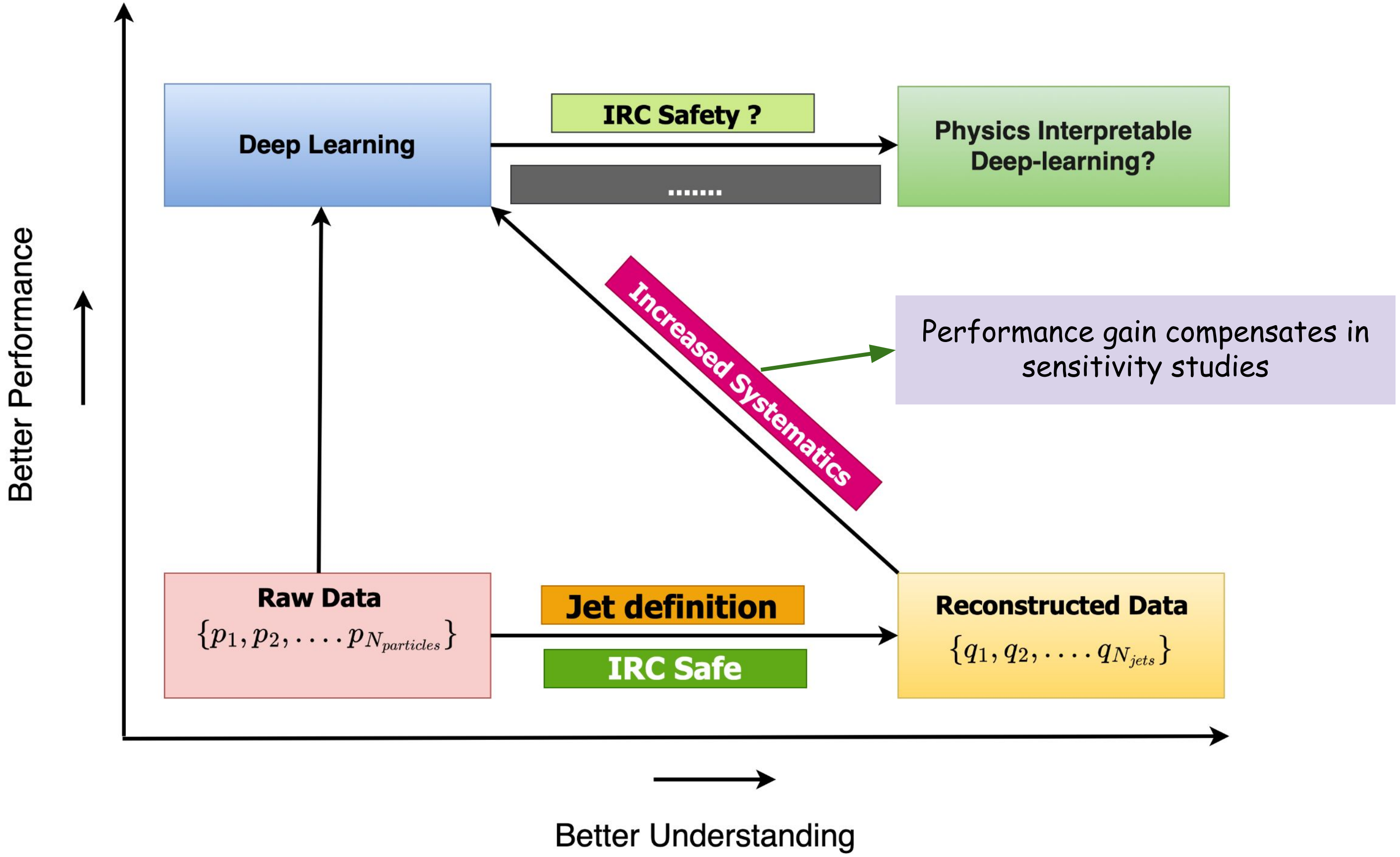
Partha Konar, ***Vishal Ngairangbam***, Michael Spannowsky

arxiv: 2109.14636 [JHEP 02 (2022) 060]

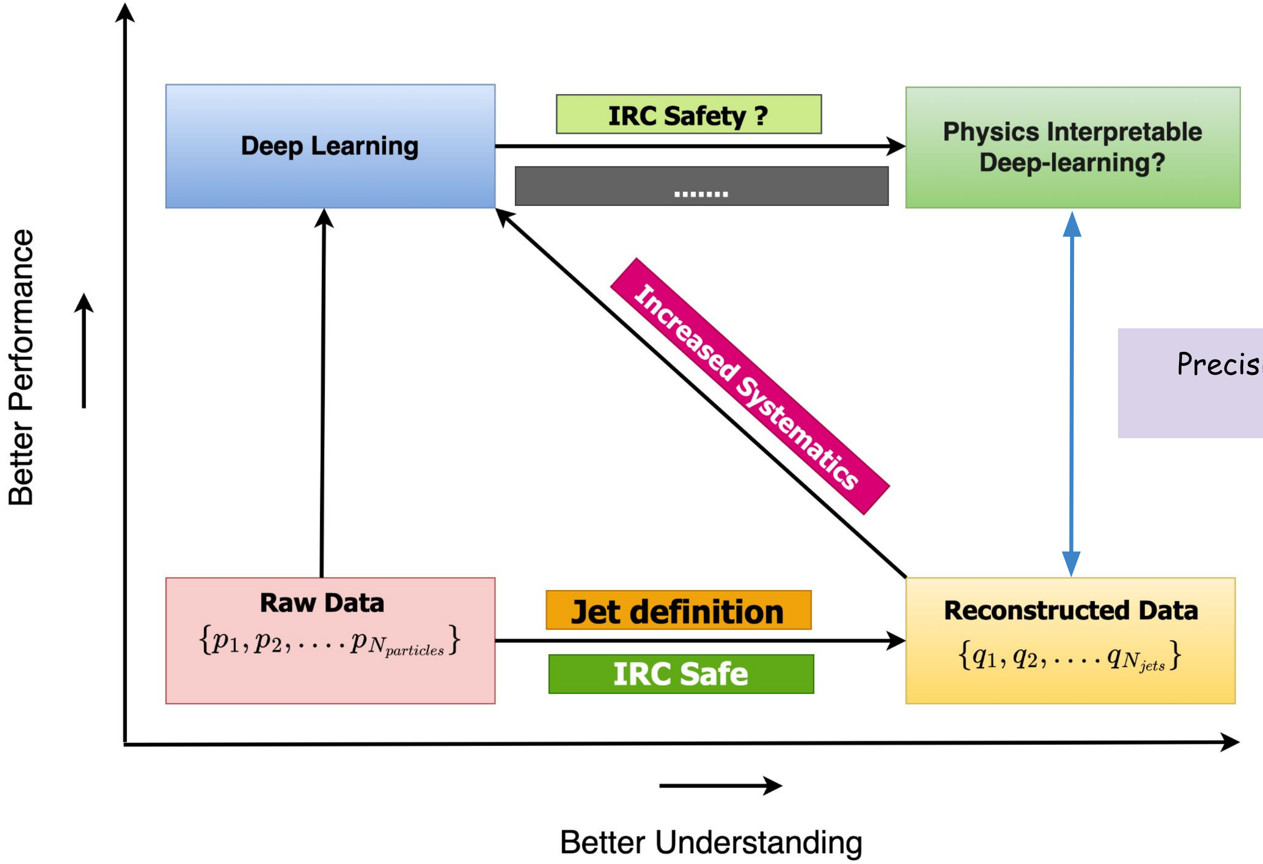
Outline

- Introduction
  - Why IRC safety in deep-learning?
  - Point clouds, graphs and message passing
- Energy-Weighted Message-passing (EPMN)
  - Building an IRC safe graph
  - IRC safe message-passing
- Results and discussions

Why Infra-red and collinear safety?



Why Infra-red and collinear safety?



Precision analyses with deep learning?

Infra-red and Collinear (IRC) Safe observables

For an observable $\mathcal{O}_n$ defined on n particles.

$$\mathcal{O}_{n+1}(p_a,.,p_b,p_r,p_s,p_c,.) \rightarrow \mathcal{O}_n(p_a,.,p_b,p_q,p_c,.)$$

In the infra-red $(z_r \rightarrow 0 \text{ or } z_s \rightarrow 0)$ or collinear limits $(\Delta_{rs} \rightarrow 0)$

Calculable in pQCD!!

For a splitting: $q \rightarrow r + s$

$$p_q = p_r + p_s$$

$$p_q = (z_q, \hat{p}_q)$$
$$p_r = (z_r, \hat{p}_r)$$
$$p_s = (z_s, \hat{p}_s)$$

Infra-red and Collinear (IRC) Safe observables

For an observable $\mathcal{O}_n$ defined on n particles.

$$\mathcal{O}_{n+1}(p_a,.,p_b,p_r,p_s,p_c,.) \rightarrow \mathcal{O}_n(p_a,.,p_b,p_q,p_c,.)$$

In the infra-red $(z_r \rightarrow 0 \text{ or } z_s \rightarrow 0)$ or collinear limits $(\Delta_{rs} \rightarrow 0)$

Calculable in pQCD!!

For a splitting: $q \rightarrow r + s$

$$p_q = p_r + p_s$$

$$p_q = (z_q, \hat{p_q})$$
$$p_r = (z_r, \hat{p_r})$$
$$p_s = (z_s, \hat{p_s})$$

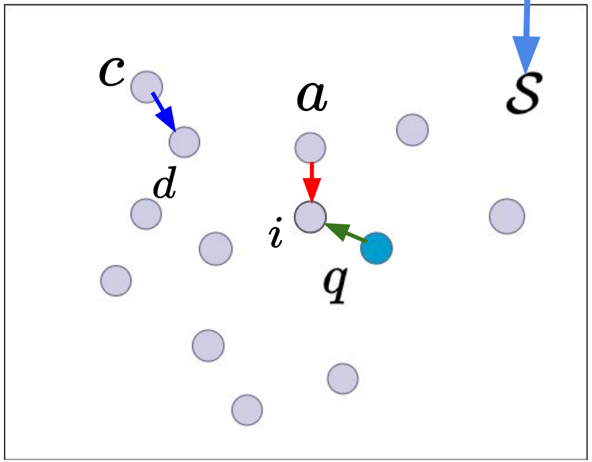
=> concentrates on information of the hard interaction

=> relatively insensitive to low-energy non-perturbative effects

Graphs: Compact efficient data structures

$\mathcal{S} = \{a, b, i, q, ..\}$

Node Set: all particles within a jet



$\mathcal{E} = \{(i,a), (i,q), (d,c), \ldots\}$

Edge set

A graph $G(\mathcal{S}, \mathcal{E})$ defined on a set $\mathcal{S}$, with edge set $\mathcal{E}$

Node-features: $\{\mathbf{h}_a, \mathbf{h}_b, \mathbf{h}_c, \ldots\}$

Four-momenta, charge, etc,

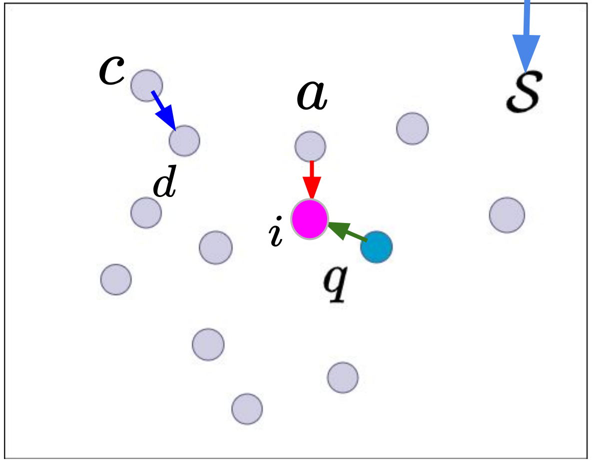
Edge-features: $\{\mathbf{e}_{ia}, \mathbf{e}_{iq}, \mathbf{e}_{dc}, \ldots\}$

$m_{ia}, \Delta R_{ia} etc,$

Message-passing neural networks

$\mathcal{S} = \{a, b, i, q, ..\}$

Node Set: all particles within a jet



$\mathcal{E} = \{(i, a), (i, q), (d, c), \ldots\}$

Edge set

Message passing operation

Message-passing (edge-centric) $\Phi(\mathbf{h}_i, \mathbf{h}_q)$

Node readout (node-centric) $\mathbf{H}_i = \sum_{j \in \mathcal{N}[i]} \Phi(\mathbf{h}_i, \mathbf{h}_j)$

$\mathcal{N}[i]$ = Set of all nodes with incoming connections to i

Graph-readout (full graph)

$$\mathbf{g} = \sum_{j \in \mathcal{S}} \mathbf{H}_j$$

To Dense network

Deep-sets vs Message-passing neural networks(MPNN)

Deep-sets per-particle map: $\Phi(p_i)$	**MPNNs** Message-function: $\Phi(p_i, p_j)$
Cannot extract inter-particle correlations	Can extract **inter-particle correlations**
Only single particle information	**Graph construction algorithm** controls information extraction at first layer(via node-readout)
Iterative application has no additional complexity on feature extraction, except functional composition $\Phi'(\Phi(p_i))$	Gradual increase in information in node-features, **after each iteration**
No such control	**Number of iterations** control the scope of information contained in the final node-feature

Energy Flow Networks : IRC safe deep-sets framework

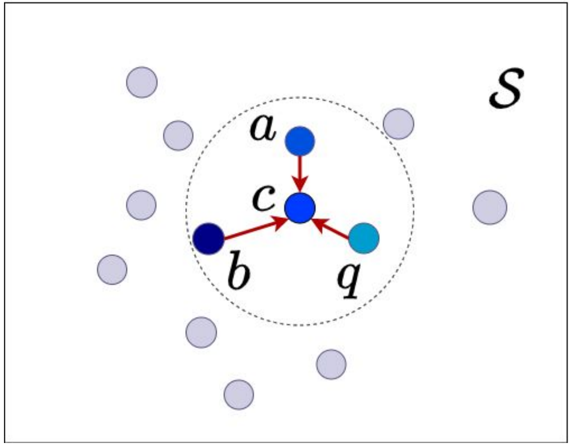
JHEP 01 (2019) 121, Komiske, Metodiev, Thaler

IRC safe Graph Construction

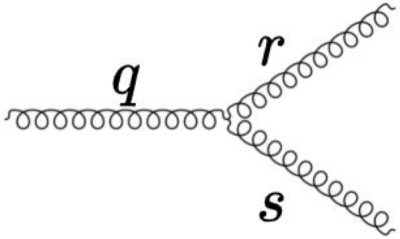
Jet-graph: k-nearest neighbour

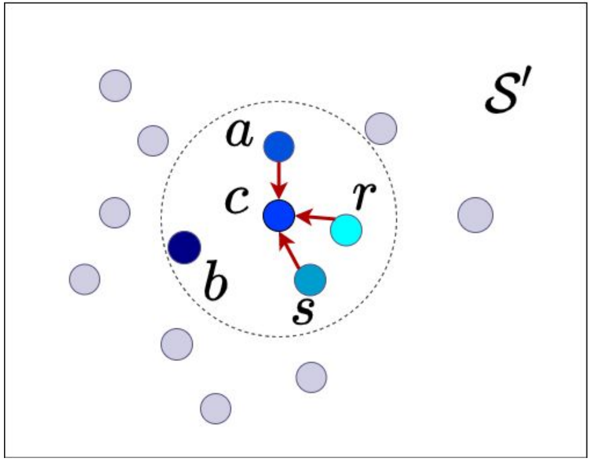
$k = 3$

(η, ϕ)-plane



k-NN graph structurally IRC unsafe!





$\mathcal{N}(c) = \{a, b, q\}$

$\mathcal{N}'(c) = \{a, r, s\}$

$$\mathbf{h}_i^{(l+1)} = \square_{j \in \mathcal{N}(i)}^{local} \, {}^i \mathbf{m}_j$$

$$\lim_{z_r \to 0} \mathbf{h}_c'^{(l+1)} \neq \mathbf{h}_c^{(l+1)}$$

Similar in the collinear limit

Node readout and IRC safety

For a splitting: $q \to r + s$

$$p_q = p_r + p_s$$

$$p_q = (z_q, \hat{p}_q)$$
$$p_r = (z_r, \hat{p}_r)$$
$$p_s = (z_s, \hat{p}_s)$$

In the infra-red $(z_r \to 0 \text{ or } z_s \to 0)$ or collinear limits $(\Delta_{rs} \to 0)$

min / max

= Selects a single node in the neighbourhood

$\Rightarrow$ C unsafe when the particular node splits

mean

= Depends explicitly on the number of nodes in $\mathcal{N}(i)$

$\Rightarrow$ Arbitrary number of emmisions in the enhanced regions (soft or collinear)

sum

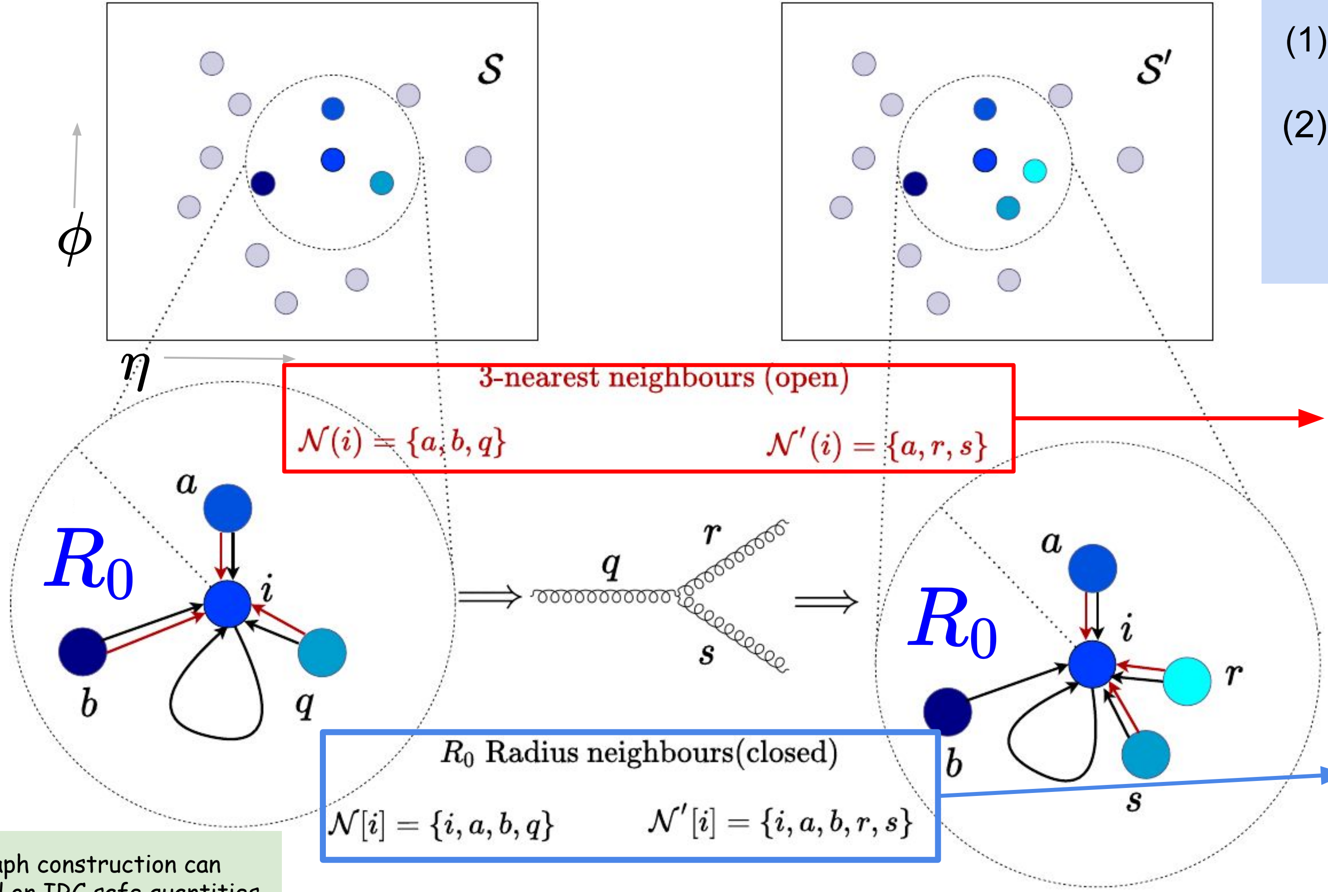
= (semi) inclusive with no explicit dependence on the number of nodes in $\mathcal{N}(i)$

semi??

$i \notin \mathcal{N}(i)$ [open neighbourhood]

$i \in \mathcal{N}[i]$ [closed neighbourhood]

IRC safe jet-graphs



(1) Total energy conserved

(2) Both r,s present in $\mathcal{N}'[i]$

if mother in $\mathcal{N}[i]$

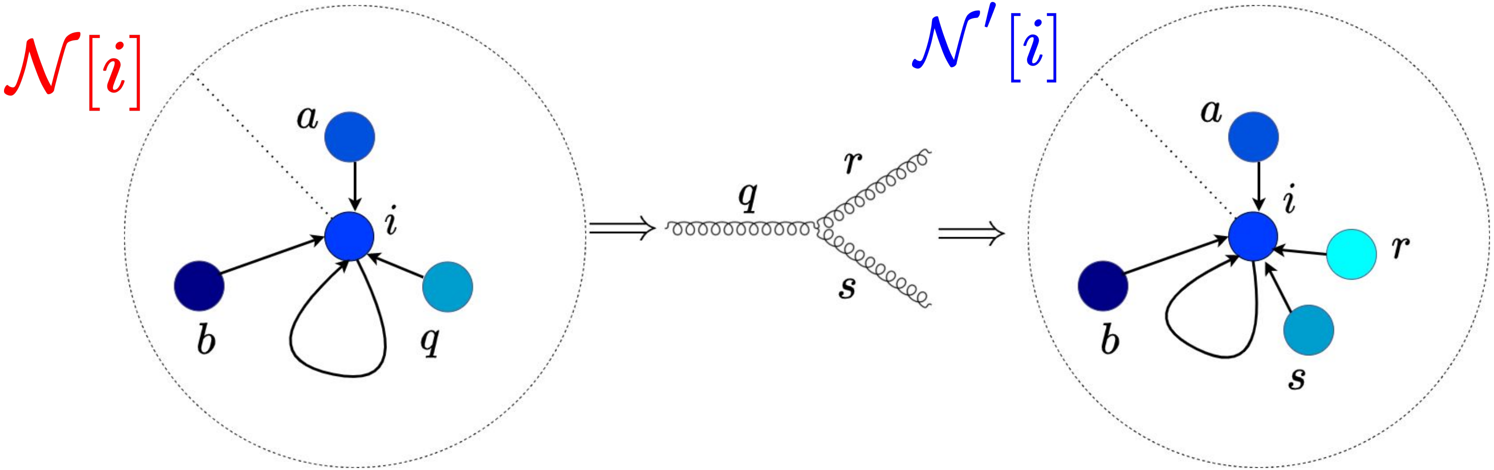
Popular choice (k-NN)
IRC unsafe!!!

IRC safe prescription!

Graph construction can depend on IRC safe quantities defined on and within the jet

Energy-weighted Message Passing Network

Energy-weighted Message-passing Network

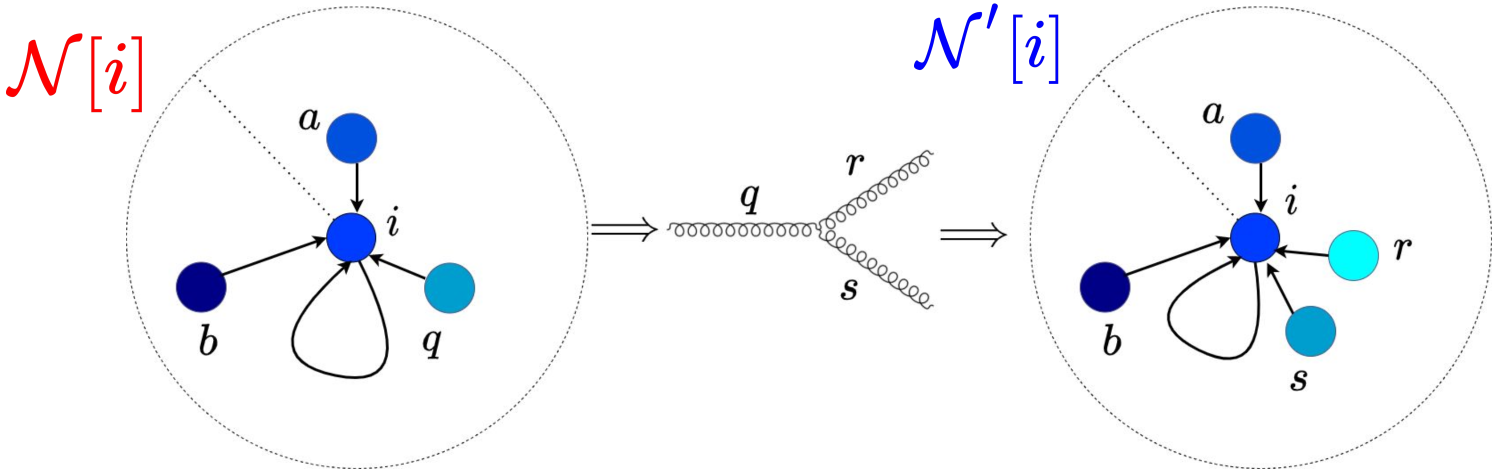


Message-passing operation

$$\mathbf{H}_i = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\Phi}(\hat{p}_i, \hat{p}_j)$$

$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

Energy-weighted Message-passing Network



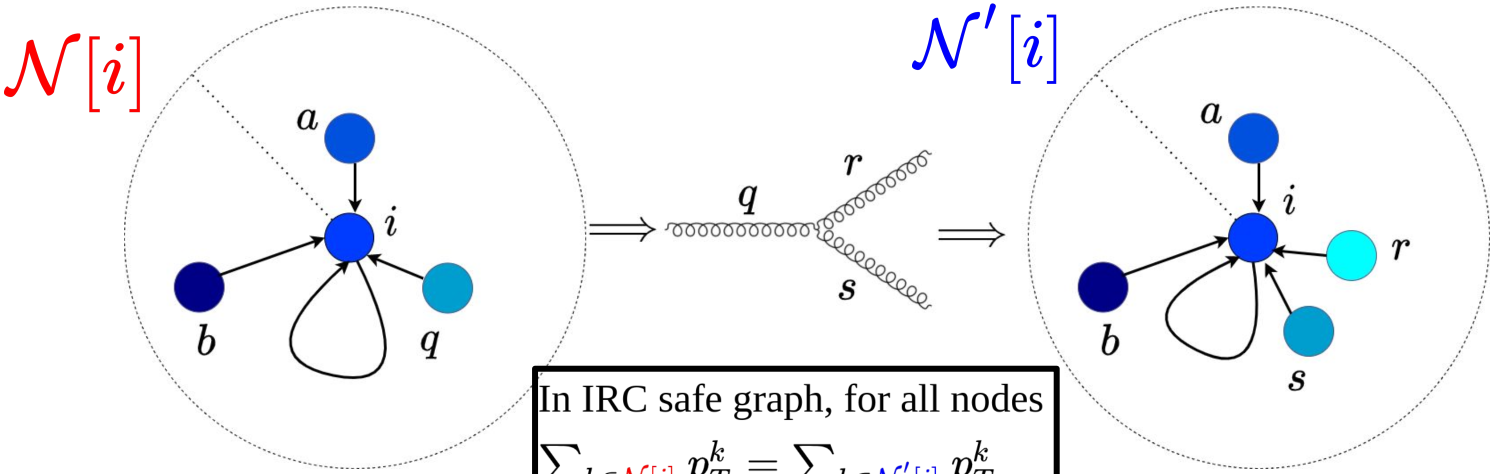
Message-passing operation

$$\mathbf{H}_i = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\Phi}(\hat{p}_i, \hat{p}_j)$$

$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

IR Safety: $z_j \to 0 \Rightarrow \omega_j^{(\mathcal{K})} \to 0$ for any $\mathcal{K}$

Energy-weighted Message-passing Network



In IRC safe graph, for all nodes

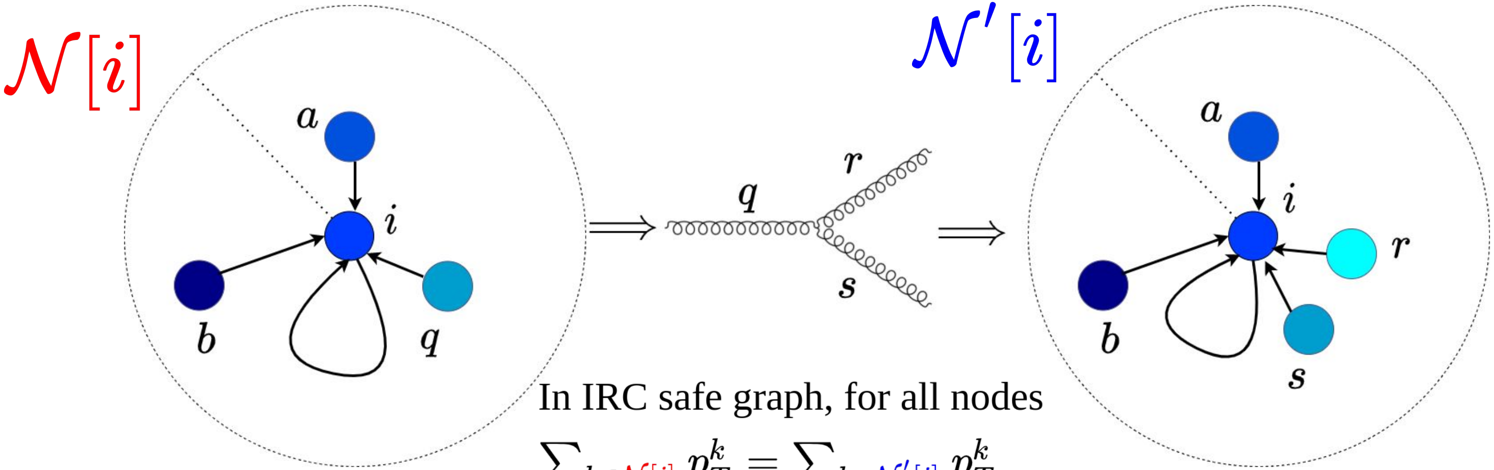
$\sum_{k\in\mathcal{N}[i]} p_T^k = \sum_{k\in\mathcal{N}'[i]} p_T^k$

Message-passing operation

$$\mathbf{H}_i = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \; \hat{\Phi}(\hat{p}_i, \hat{p}_j)$$

$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

Energy-weighted Message-passing Network



In IRC safe graph, for all nodes

$$\sum_{k\in\mathcal{N}[i]} p_T^k = \sum_{k\in\mathcal{N}'[i]} p_T^k$$

$$p_T^q = p_T^r + p_T^s \Rightarrow \omega_q^{(\mathcal{N}[i])} = \omega_r^{(\mathcal{N}'[i])} + \omega_s^{(\mathcal{N}'[i])}$$

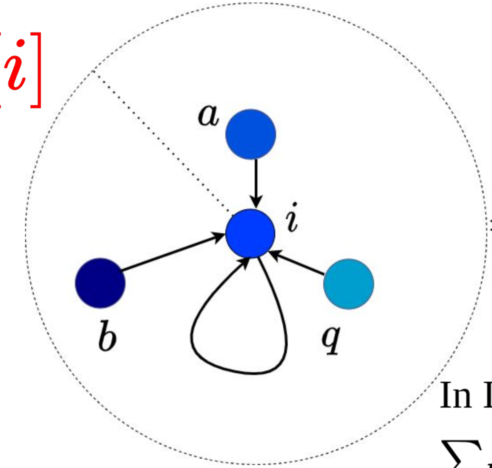
Message-passing operation

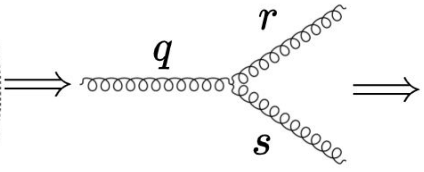
$$\mathbf{H}_i = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\Phi}(\hat{p}_i, \hat{p}_j)$$

$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

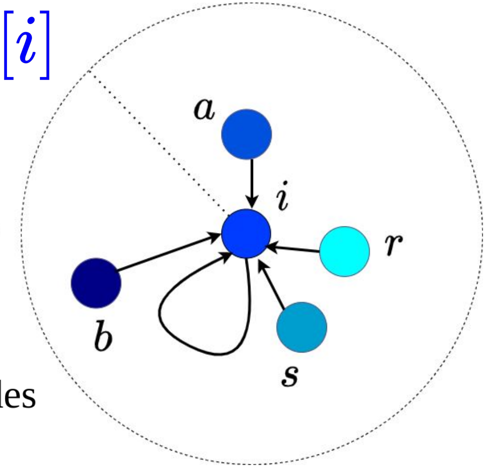
Energy-weighted Message-passing Network

$\mathcal{N}[i]$





$\mathcal{N}'[i]$



In IRC safe graph, for all nodes

$$\sum_{k\in\mathcal{N}[i]} p_T^k = \sum_{k\in\mathcal{N}'[i]} p_T^k$$

$$p_T^q = p_T^r + p_T^s \Rightarrow \omega_q^{(\mathcal{N}[i])} = \omega_r^{(\mathcal{N}'[i])} + \omega_s^{(\mathcal{N}'[i])}$$

Message-passing operation

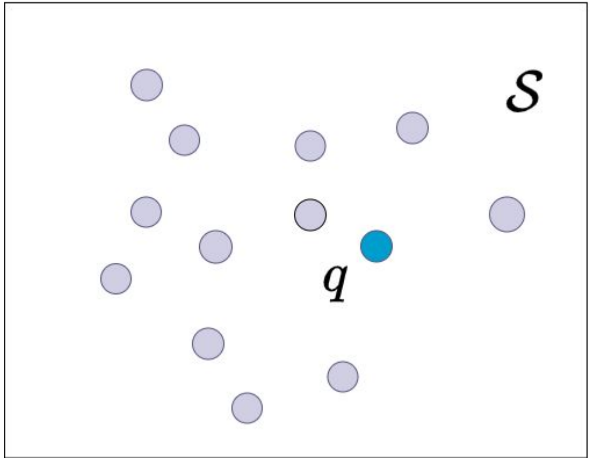
$$\mathbf{H}_i = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \, \hat{\Phi}(\hat{p}_i, \hat{p}_j)$$

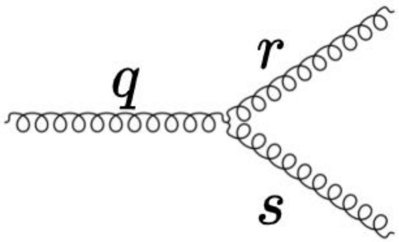
$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

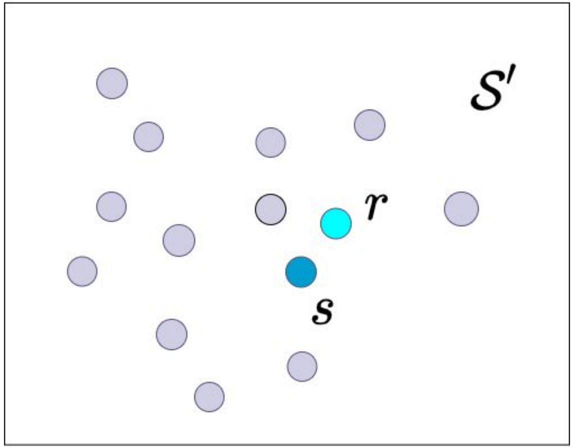
C Safety: $\hat{p}_q = \hat{p}_r = \hat{p}_s$

$$\omega^{(\mathcal{N}[i])} \, \hat{\Phi}(\hat{p}_i, \hat{p}_q) = \omega^{(\mathcal{N}'[i])} \, \hat{\Phi}(\hat{p}_i, \hat{p}_r) + \omega^{(\mathcal{N}[i])} \, \hat{\Phi}(\hat{p}_i, \hat{p}_s)$$

EMPN: Iterative application



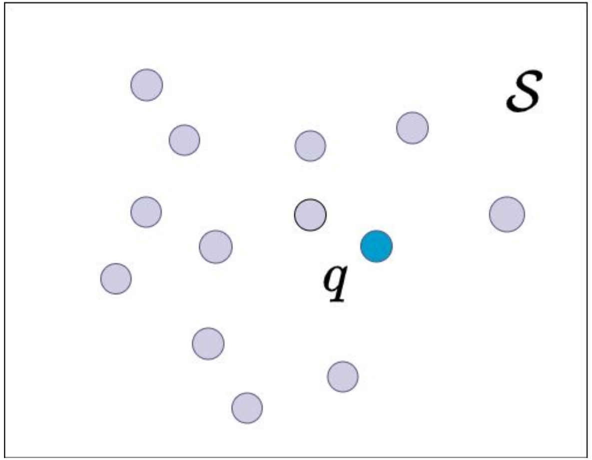


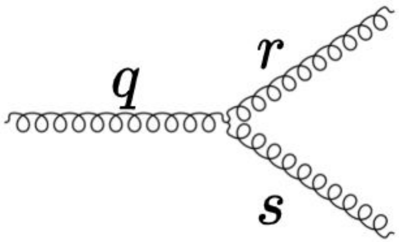


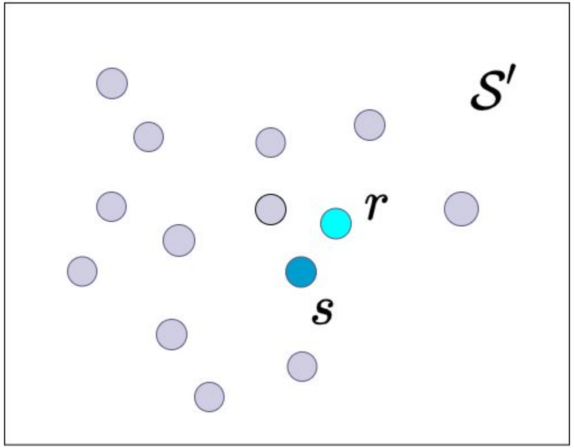
C limit:

$$\mathbf{h}_q^{(1)} = \mathbf{h}_r^{(1)} = \mathbf{h}_s^{(1)}$$

EMPN: Iterative application





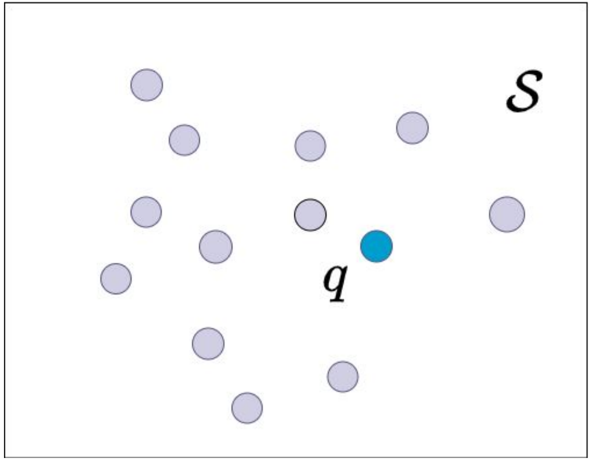


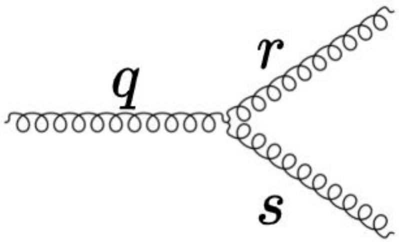
C limit:

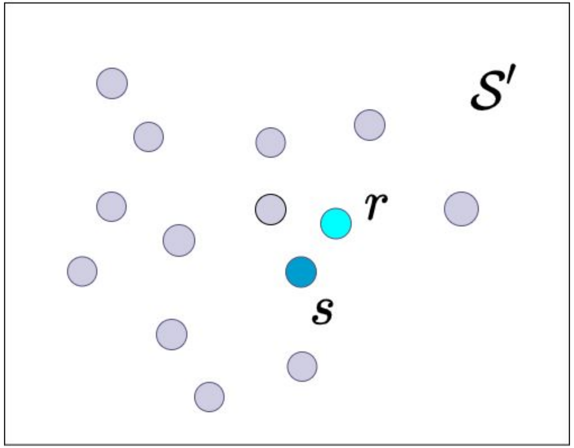
$$\mathbf{h}_q^{(1)} = \mathbf{h}_r^{(1)} = \mathbf{h}_s^{(1)}$$

$$\mathbf{h}_i^{(2)} = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])}\,\hat{\mathbf{\Phi}}^{(1)}(\mathbf{h}_i^{(1)}, \mathbf{h}_j^{(1)})$$

EMPN: Iterative application







Iterative Application is possible!

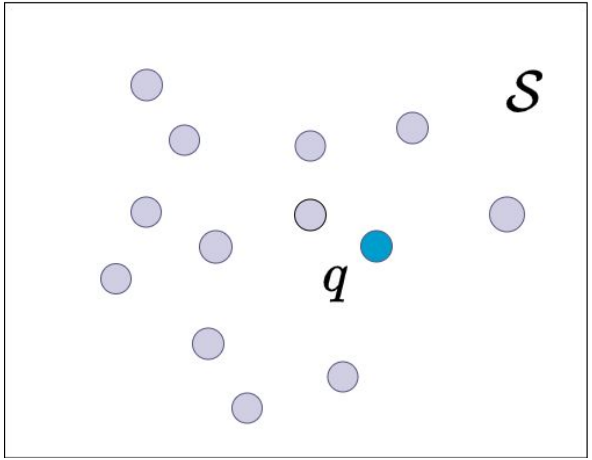
C limit:

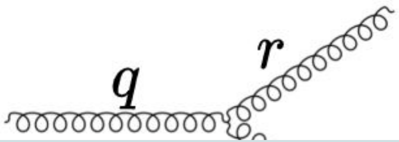
$$\mathbf{h}_i^{(l+1)} = \sum_{j\in\mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\Phi}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)})$$

$$\mathbf{h}_q^{(l)} = \mathbf{h}_r^{(l)} = \mathbf{h}_s^{(l)}$$

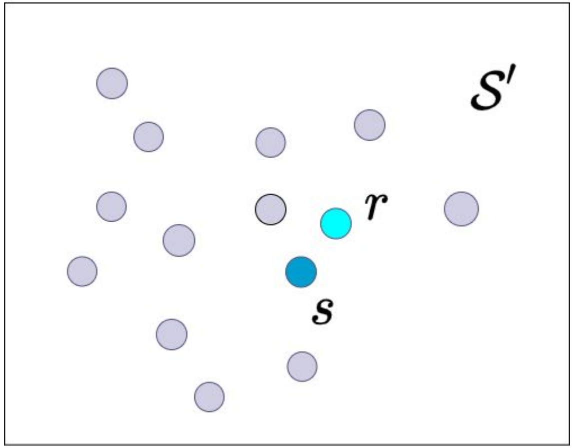
$$\mathbf{h}_i^{(0)} = \hat{p}_i$$

EMPN: Iterative application





Can be radius neighbourhood in the space of $\mathrm{h}_i^{(l)}$
=> graph can be dynamic!



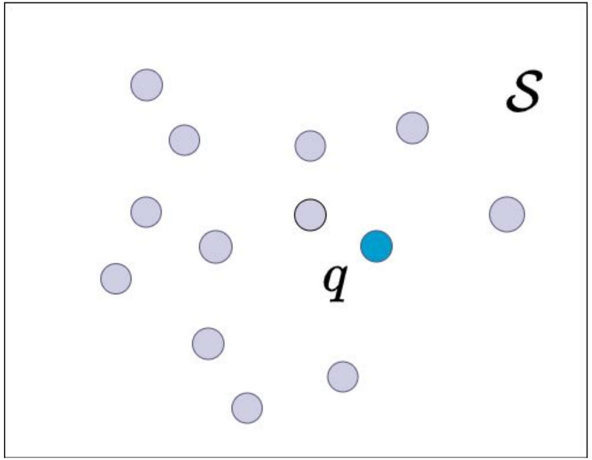
C limit:

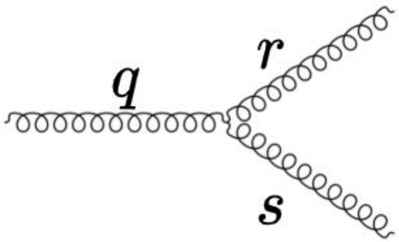
$$\mathbf{h}_q^{(l)} = \mathbf{h}_r^{(l)} = \mathbf{h}_s^{(l)}$$

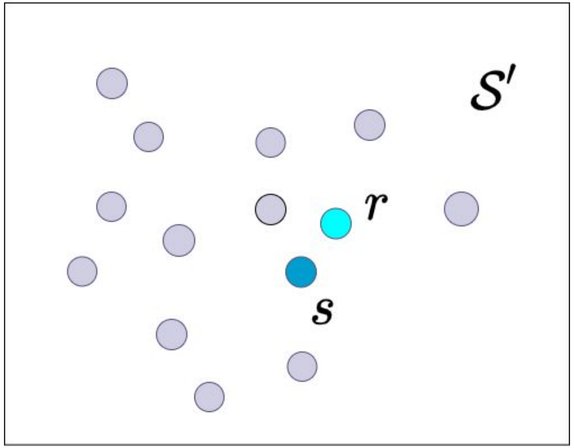
$$\mathbf{h}_i^{(l+1)} = \sum_{j \in \mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\boldsymbol{\Phi}}^{(l)} (\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)})$$

$$\mathbf{h}_i^{(0)} = \hat{p}_i$$

EMPN: Graph-readout







C limit:

$$\mathbf{h}_q^{(L)} = \mathbf{h}_r^{(L)} = \mathbf{h}_s^{(L)}$$

$L =$ num. iterations

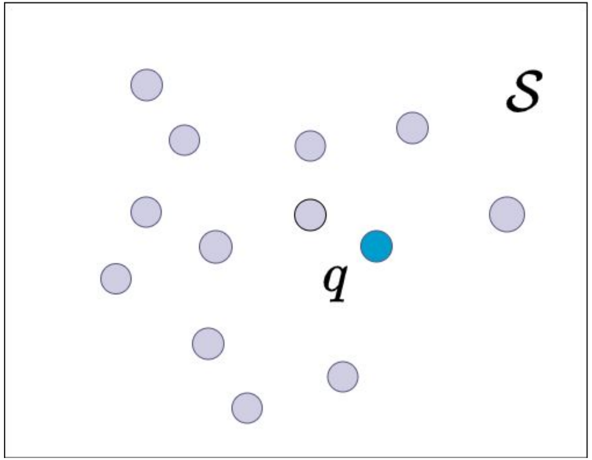
Soft or collinear nodes have a node representation even in the IRC limit!!!

=> IRC safety cannot be defined on a node representation!

=> need to define an IRC safe graph representation

Graph on $\mathcal{S}'$ contain $\mathbf{h}_r^{(L)}$ and $\mathbf{h}_s^{(L)}$

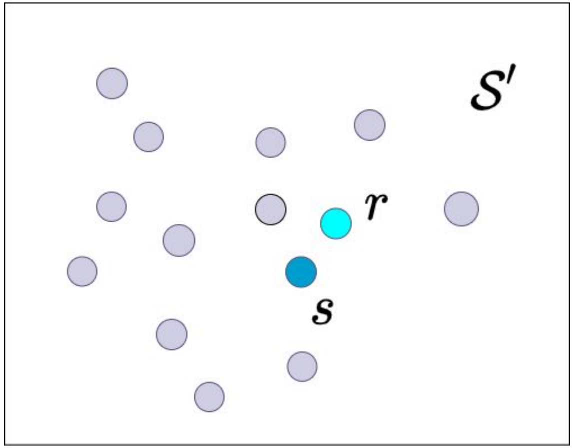
EMPN: Graph-readout



Graph readout

$$\mathbf{g} = \sum_{j \in \mathcal{S}} \omega_j^{(\mathcal{S})} \, \mathbf{h}_j^{(L)}$$

Representation of the full jet is IRC safe



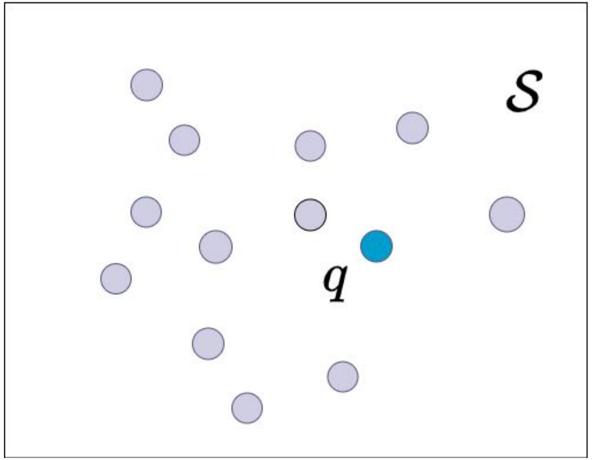
C limit:

$$\mathbf{h}_q^{(L)} = \mathbf{h}_r^{(L)} = \mathbf{h}_s^{(L)}$$

$L =$ num. iterations

$$\omega_i^{(\mathcal{S})} = z_i$$

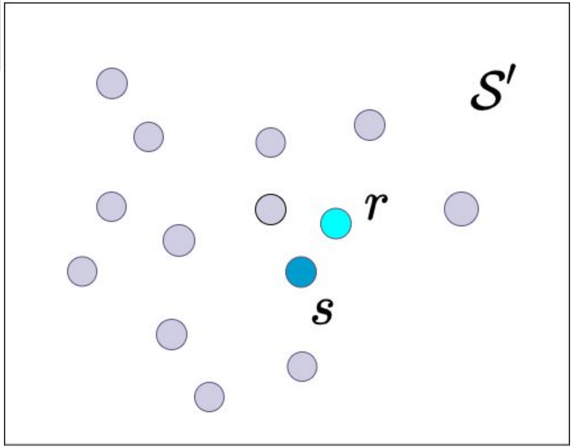
EMPN: Graph-readout



Information content determined by R_0 and L

Graph readout

$$\mathbf{g} = \sum_{j\in\mathcal{S}} \omega_j^{(\mathcal{S})} \mathbf{h}_j^{(L)}$$



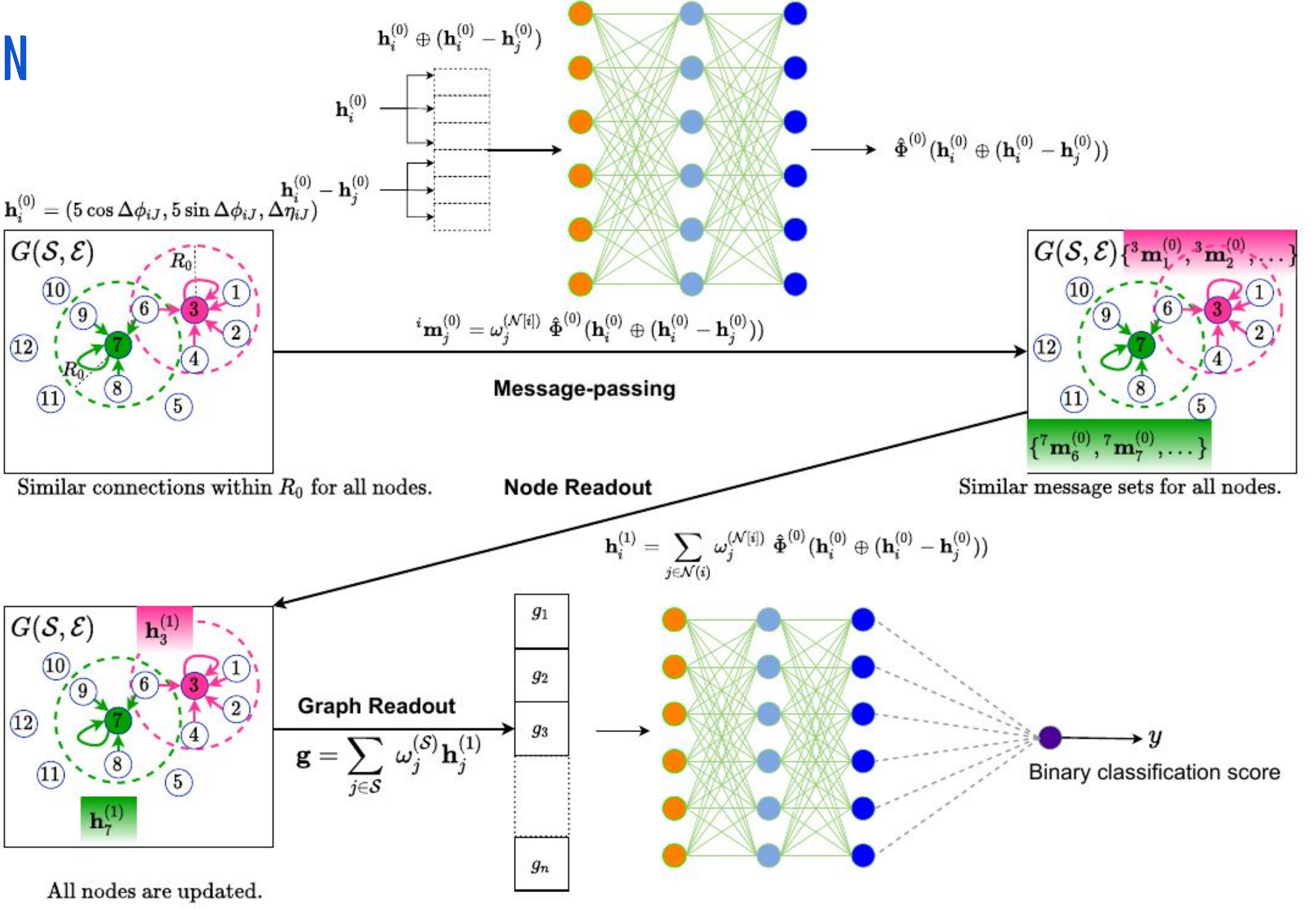
C limit:

$$\mathbf{h}_q^{(L)} = \mathbf{h}_r^{(L)} = \mathbf{h}_s^{(L)}$$

$L =$ num. iterations

$$\omega_i^{(\mathcal{S})} = z_i$$

EMPN



$\mathbf{h}_i^{(0)} \oplus (\mathbf{h}_i^{(0)} - \mathbf{h}_j^{(0)})$
$\mathbf{h}_i^{(0)}$
$\mathbf{h}_i^{(0)} - \mathbf{h}_j^{(0)}$
$\hat{\Phi}^{(0)}(\mathbf{h}_i^{(0)} \oplus (\mathbf{h}_i^{(0)} - \mathbf{h}_j^{(0)}))$
$\mathbf{h}_i^{(0)} = (5\cos\Delta\phi_{iJ}, 5\sin\Delta\phi_{iJ}, \Delta\eta_{iJ})$
$G(\mathcal{S}, \mathcal{E})$
R_0
${}^i\mathbf{m}_j^{(0)} = \omega_j^{(\mathcal{N}[i])}\, \hat{\Phi}^{(0)}(\mathbf{h}_i^{(0)} \oplus (\mathbf{h}_i^{(0)} - \mathbf{h}_j^{(0)}))$
Message-passing
$G(\mathcal{S}, \mathcal{E})\{{}^3\mathbf{m}_1^{(0)}, {}^3\mathbf{m}_2^{(0)}, \dots\}$
$\{{}^7\mathbf{m}_6^{(0)}, {}^7\mathbf{m}_7^{(0)}, \dots\}$
Similar connections within R_0 for all nodes.
Node Readout
Similar message sets for all nodes.
$\mathbf{h}_i^{(1)} = \sum_{j\in\mathcal{N}(i)} \omega_j^{(\mathcal{N}[i])}\, \hat{\Phi}^{(0)}(\mathbf{h}_i^{(0)} \oplus (\mathbf{h}_i^{(0)} - \mathbf{h}_j^{(0)}))$
$G(\mathcal{S}, \mathcal{E})$
$\mathbf{h}_3^{(1)}$
$\mathbf{h}_7^{(1)}$
Graph Readout
$\mathbf{g} = \sum_{j\in\mathcal{S}} \omega_j^{(\mathcal{S})} \mathbf{h}_j^{(1)}$
g_1
g_2
g_3
g_n
y
Binary classification score
All nodes are updated.

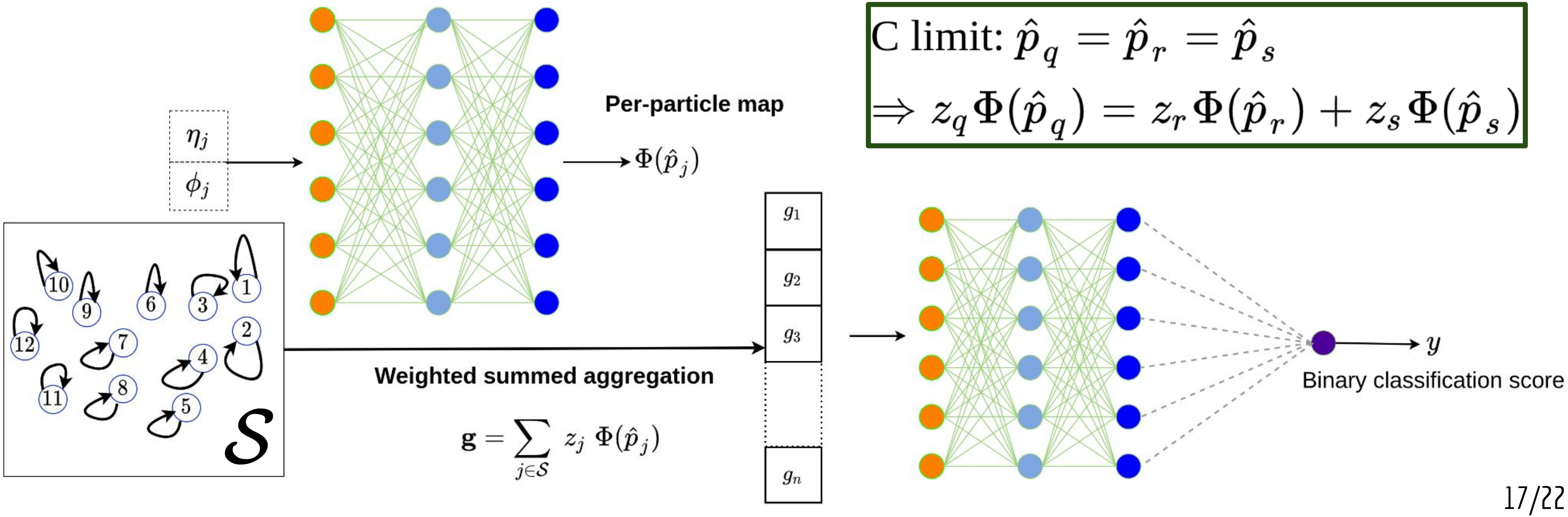
Generalising Energy Flow Networks (EFNs)

JHEP 01 (2019) 121, Komiske, Metodiev, Thaler

Per-particle map is a special message function constant for the second argument!

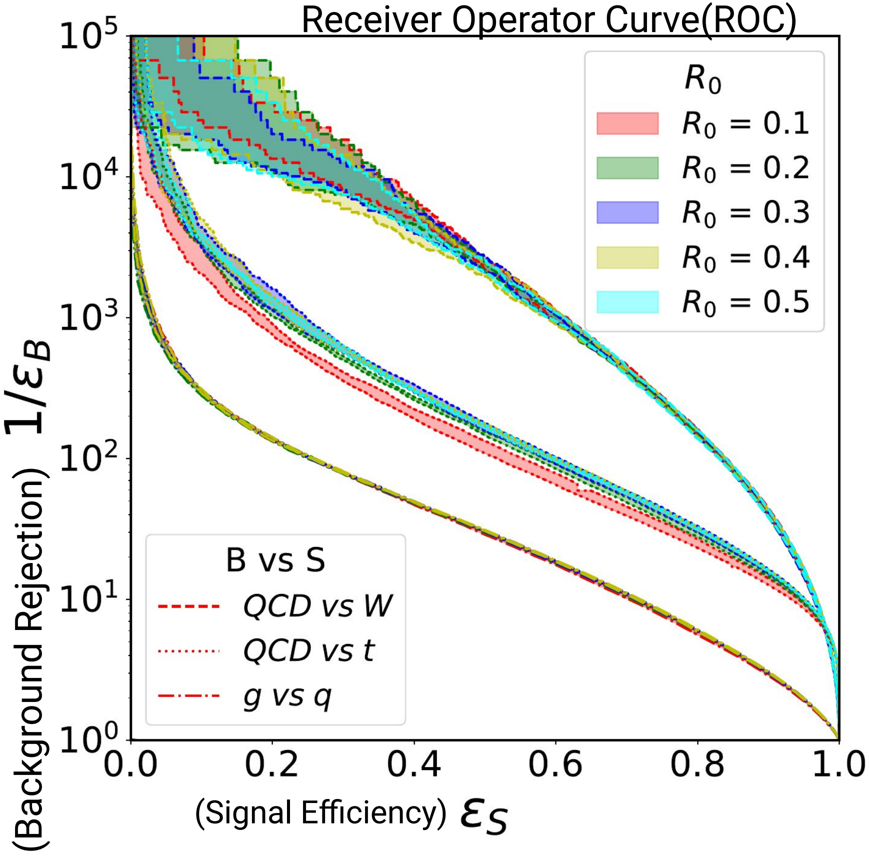
$$\text{IR limit: } z_r = 0 \Rightarrow z_r\, \Phi(\hat{p}_r) = 0$$

$$\text{C limit: } \hat{p}_q = \hat{p}_r = \hat{p}_s$$
$$\Rightarrow z_q \Phi(\hat{p}_q) = z_r \Phi(\hat{p}_r) + z_s \Phi(\hat{p}_s)$$

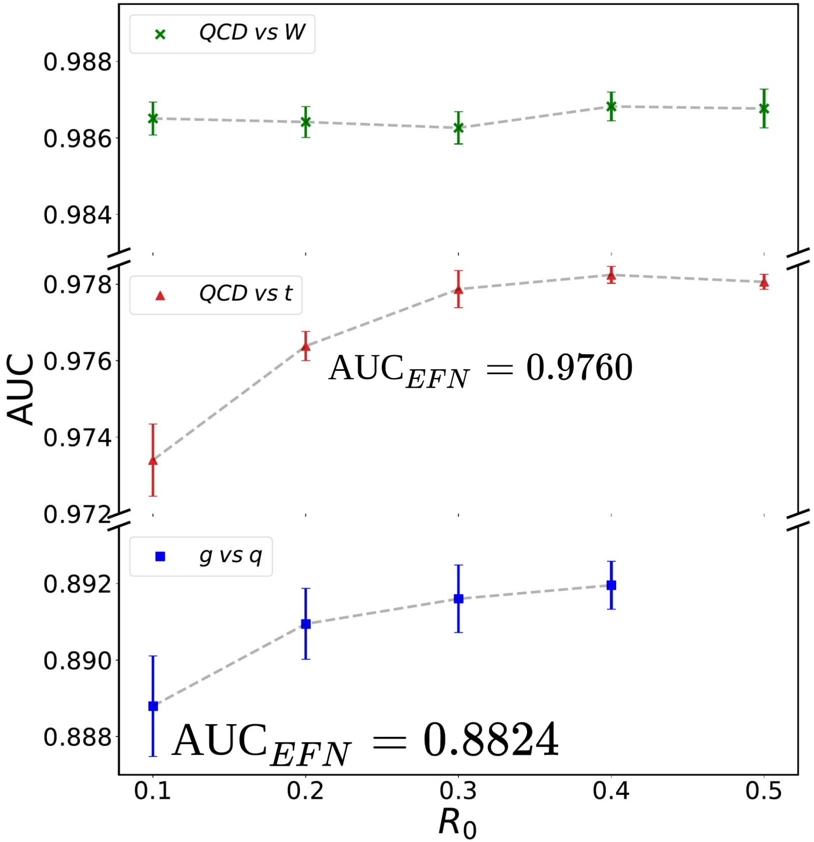


Results

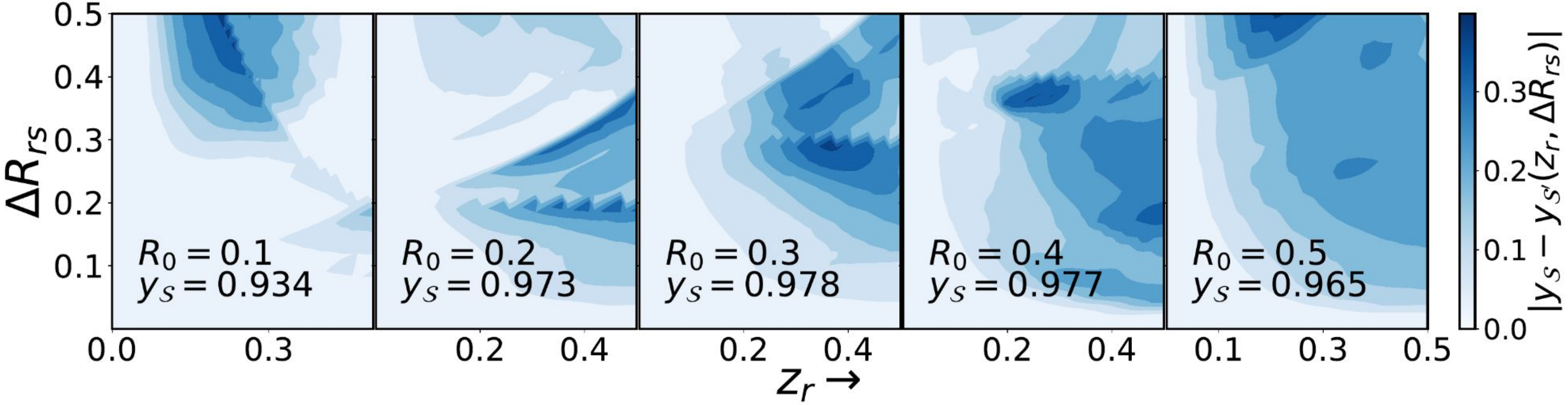
Network Performance (L=1)



Area-under(ROC)-curve(AUC) (ϵ_B, ϵ_S)



Examining IRC Safety (L=1)



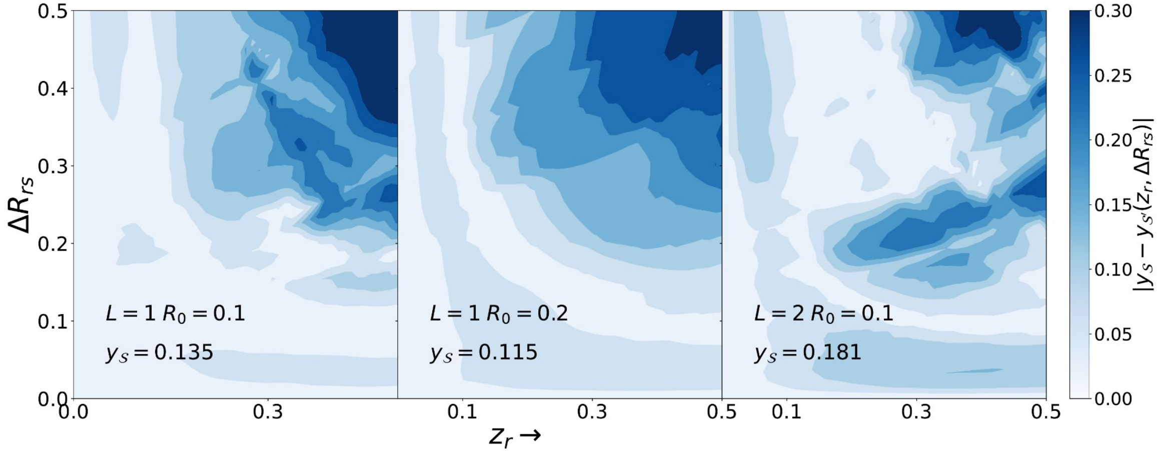
Split the hardest constituent in a jet and vary z_r and ΔR_{rs}

Network Output: $y_S \quad y_{S'}(z_r, \Delta R_{rs})$

Increasing R_0 **decreases stability** of network output to **additional emissions**

Iterative application (L=1 vs L=2) (Quark vs Gluon)

R_0	No. message passing (L)	AUC
0.1	1	0.8888 ± 0.0013
0.2	1	0.8909 ± 0.0009
0.4	1	0.8919 ± 0.0006
0.1	2	0.8932 ± 0.0006



Conclusions

- Generalised Energy Flow Networks (EFNs) to extract local correlations via message-passing operations
- Iterative application does not spoil IRC safety, performs better with reduced (?) sensitivity to soft and collinear emissions
- Devised generic graph construction algorithms which give invariant graph structure in the ***deletion*** of a soft or collinear vertex
- Possibility to structure graphs and networks with highly intuitive physics input
- General enough to study event shapes
- Can we understand the extracted features within pQCD?
- Higher point correlations with IRC safe hypergraphs?

Conclusions

- Generalised Energy Flow Networks (EFNs) to extract local correlations via message-passing operations
- Iterative application does not spoil IRC safety, performs better with reduced (?) sensitivity to soft and collinear emissions
- Devised generic graph construction algorithms which give invariant graph structure in the ***deletion*** of a soft or collinear vertex
- Possibility to structure graphs and networks with highly intuitive physics input
- General enough to study event shapes
- Can we understand the extracted features within pQCD?
- Higher point correlations with IRC safe hypergraphs?

THANK YOU

Dataset Details

Sl. No	Jet Class	Parton-level	MPI	Detector Simulation (Delphes3)	Jet Radius (anti-kT)	Transverse momentum [GeV]	Classification Scenario
1.	Gluon	Pythia8	Yes	No	0.4	[500,550]	Gluon vs Quark
2.	Quark	Pythia8	Yes	No	0.4	[500,550]	Gluon vs Quark
3.	QCD	Pythia8	No	Yes	0.8	[550,650]	QCD vs Top/W
4.	Top	Pythia8	No	Yes	0.8	[550,650]	QCD vs Top
5.	W	Madgraph5	No	Yes	0.8	[550,650]	QCD vs W

[1-2] Publicly available q/g dataset [Komiske et.al] (used in EFNs)

[3,4] Publicly available top tagging dataset [Kasieczka et.al] (used in EFNs)

[5] Generated with same specifications as [3,4]