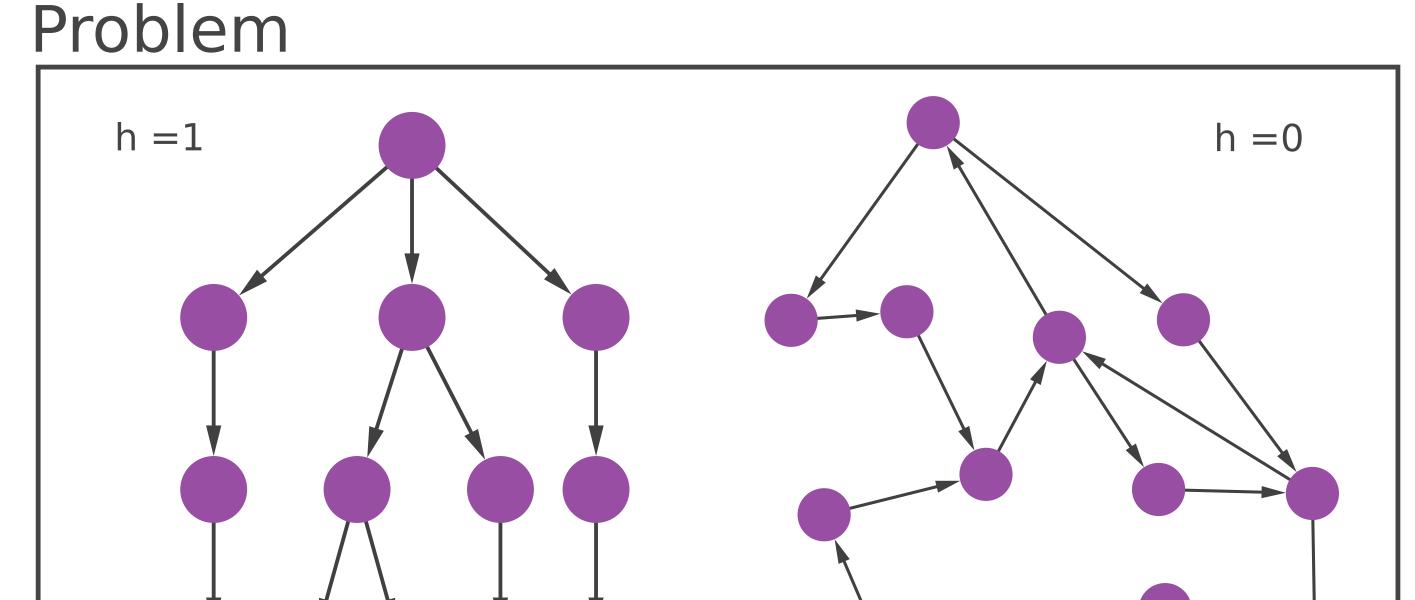
Using Arborescences to Estimate Hierarchicalness in Directed Complex Networks

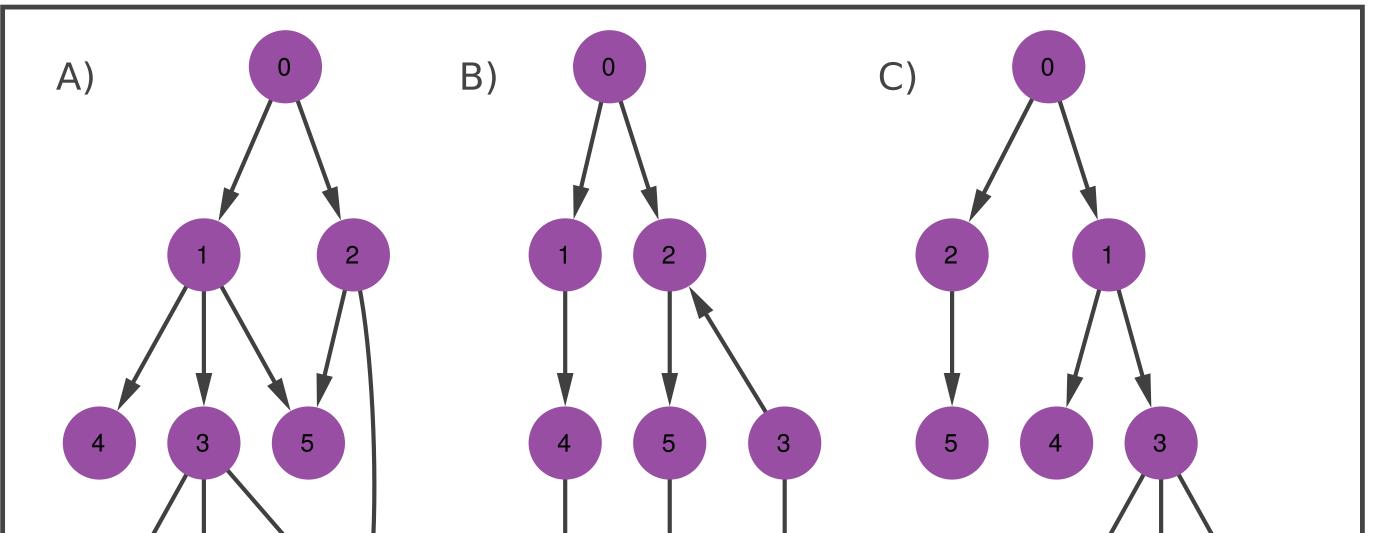
Michele Coscia

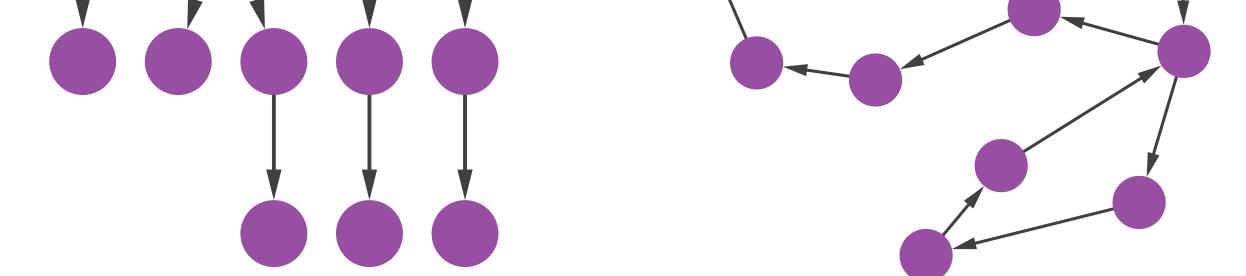
Center for international Development, Harvard University

michele_coscia@hks.harvard.edu



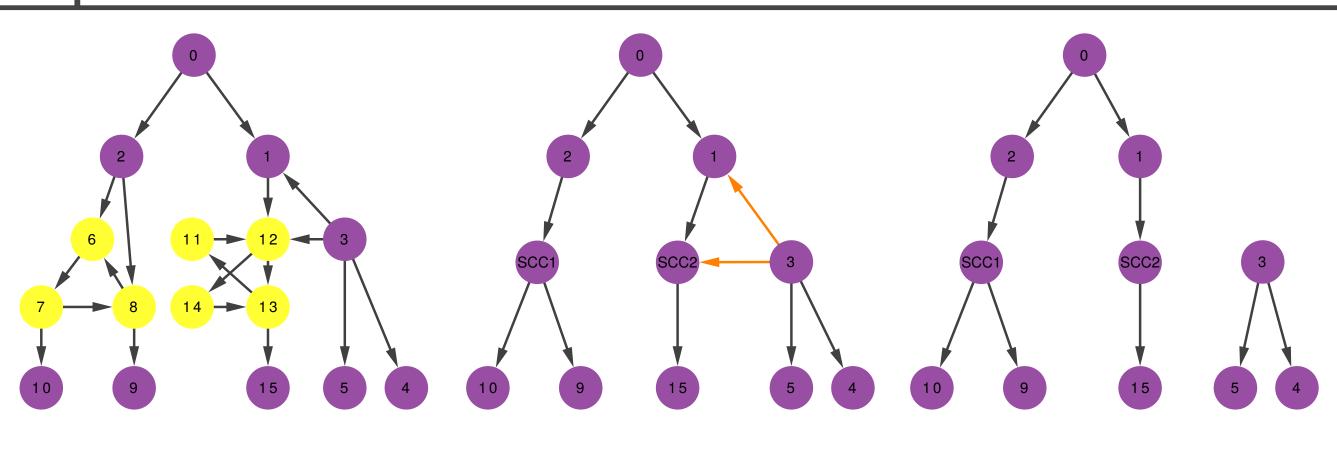
Definitions





Given a directed graph G = (V, E) return a score h in [0, 1] such that h = 0 for a graph with no hierarchical structure and h = 1 for a graph with a perfect hierarchy.

Implementation

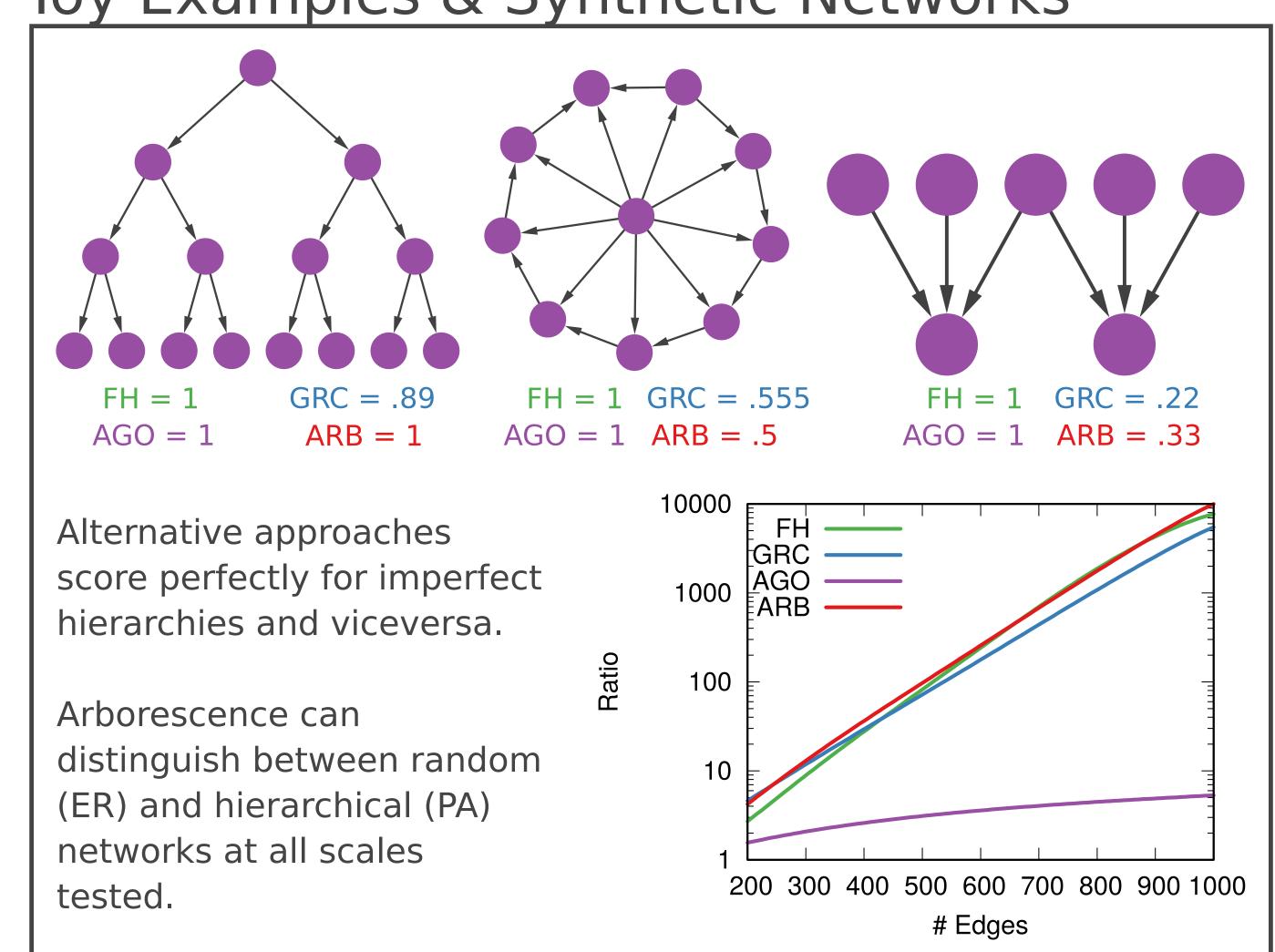


Identify strongly connected components and collapse them into a single node (highlighted in yellow). This makes the network into a directed acyclic graph.

A) Directed Acyclic Graph: directed graph with no strongly connected components.

B) Directed Tree: directed acyclic graph with no strongly connected components even if we ignore the edge direction.

C) Arborescence: a directed tree in which all nodes have in-degree of one, except the root which has in-degree of zero.



Toy Examples & Synthetic Networks

Identify edges going against the hierarchy flow and remove them (highlighted in orange).

The resulting network is an arborescence (forest) -- on the right.

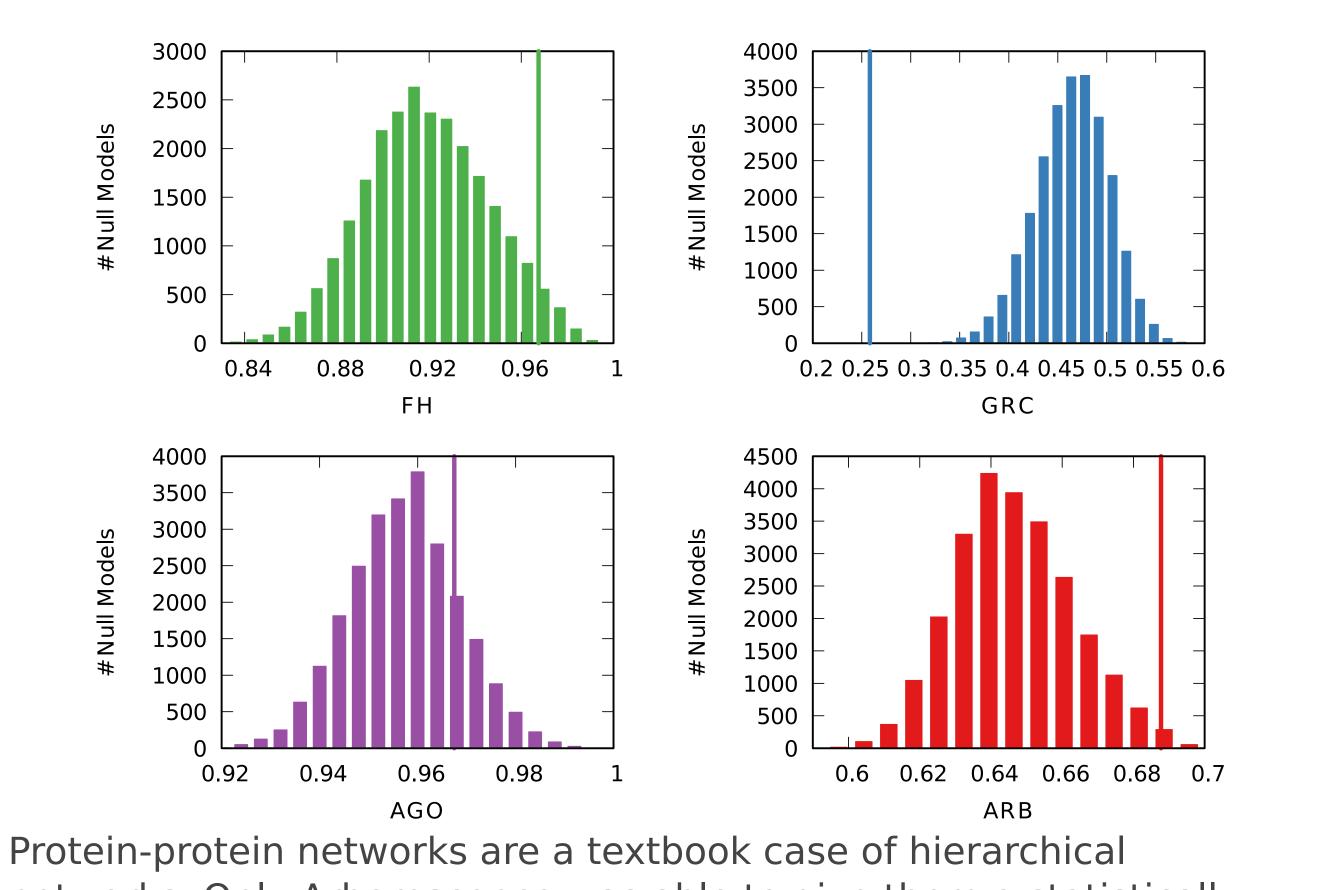
The Arborescence score is the number of edges surviving in the final step over the original network's number of edges.

In the example: h = 9 / 20 = .45

Literature Agreement

Network	V	E	E^{*}	t	FH	GRC	AGO	ARB	Agreement?
Colombia Social	863	$9,\!639$	131	0.021	_	×	_	_	\checkmark
Colombia Mobility	863	$6,\!614$	22	0.012	×	×	×	×	\checkmark
UN Migration Stocks*	175	$1,\!546$	8	0.002	×	×	×	×	\checkmark
O*Net*	496	$2,\!848$	14	0.005	×	×	×	×	\checkmark
C. Elegans Frontal	131	764	14	0.002	+	×	+	×	
Hiring Business	113	$3,\!515$	40	0.005	+	+	+	+	\checkmark
Hiring CS	206	$3,\!053$	77	0.006	+	+	+	+	\checkmark
Hiring History	145	$2,\!496$	60	0.004	+	+	+	+	\checkmark
Literature Citation	118	613	613	0.012	+	+	+	+	\checkmark
Literature Criticism	35	81	79	0.002	×	×	×	×	\checkmark
Physician Trust	241	$1,\!098$	71	0.004	+	+	×	+	\checkmark
Foodweb Everglades	69	916	26	0.002	+	+	+	+	\checkmark
Foodweb Maspalomas	24	82	37	0.001	+	+	+	+	\checkmark
Foodweb Chesapeake	39	177	60	0.001	×	+	+	+	\checkmark
Foodweb ChesLower	37	178	34	0.001	×	+	+	+	\checkmark
Foodweb ChesMiddle	37	209	21	0.001	×	+	+	+	\checkmark
Foodweb ChesUpper	37	215	26	0.001	+	+	+	+	\checkmark
Foodweb ChrystalC	24	125	7	0.001	×	×	+	×	\checkmark
Foodweb ChrystalD	24	100	12	0.000	×	×	×	×	\checkmark
Foodweb Mondego	46	400	27	0.001	+	+	+	+	\checkmark
Foodweb StMarks	54	356	76	0.002	×	+	+	+	\checkmark
Foodweb Michigan	39	221	17	0.001	+	+	+	+	\checkmark
Foodweb Rhode	19	53	24	0.000	+	×	×	×	\checkmark
Foodweb Florida	128	$2,\!106$	82	0.004	+	+	+	+	\checkmark
Foodweb Narragan	35	220	10	0.001	+	+	+	+	\checkmark
Yeast	1,870	$2,\!277$	$2,\!203$	0.051	×	—	×	+	





networks. Only Arborescence was able to give them a statistically significant score when compared to degree-preserving null models.