Synthesis

Current Approach

- directed (asym), but local (no paths) metric
- should be able to use a 1,2 step scheme to begin recursion for arbit complex gates
- need to integrate back single edges, such that explore pairwise and triplet (deg2) corr's
  - info or H(Y|X) def descriptive for the final target
  - but otherwise ofc full info relative to its own output, yet adds nothing...
- co-op & comp for both nodes and edges as a recursion?
  - poss alternative to testing all combos
  - along with local only resolution approach

- want to use info to describe expected # iters to resolve
  - if H(Y|X) == 1: takes at most 1 iter to reach correctness (deterministic case)
- only max P(Y|X) in fwd direction?
  - then min H(Y|X) as first recursion?
  - or normz I by H(Y)
  - unclear how to derive pid from this
- normz by op entropy, H(in_ops) (sort of a recursion)
  - normz if want metric for 'expected time it would have taken to arrive here'
  - unless also want to include potential difficulty to adjust if needed
  - or if actually processing such an op reqs more mem or time
- relation between combinations and entropy??
- pid with immed output is better for distributed approach

On Rare Events

- typical vs rare: as degree increases, #ordered sets are exp >> #unordered sets
  - unlike error, info can describe simplicity as many ordered combos per unordered combo
  - allows for more symbolic expressivity
  - at risk of loss of information/confusion/tangling
- parallel decomp as instances are viewed, with rare relevant events on minority of decomposed subsets leading the way?

- Want: min avg run time, atypical, exp rare problems are exp longer to run (throw off the avg) ?
- Need more mathm description of in what way atypical problems are exp longer to run.
- relation between this exp avg (during info) for rare events and Jarz?
- Base problem (correctness) is evaluated multiplicatively as info, in order to achieve an average over the expected set of base problems.
- mult noise as feedback

- Think in terms of Prigogine's statistical averages. Instead of over time --> converging, think as over space (large N) converging.
- PARITY as furthest from majority x determines y case. Resistant to description via an average.
- For Prigogine, simple case of MA info is greater than sum of mi info is demonstrated via a chaotic function that rapidly approaches its average. Convergence in this case is a naive solution (avg) solving --> 1 instances in the limit of many (static) instances. For instance, AND is true if the average of x = 1, false if < 1. XOR lacks such a description. This also begins to touch on why rare events should cause longer run times, but does not clearly justify exp longer run times. Exp weighting is done by Info.
- Prigogine was talking in terms of pr density. Parity case involves rapidly fluctuating y, relative to incremental changes in input density (one x_i). Chaos with clear MA as fluctuations in x, but predictable resulting y.

- Earlier idea that complexity is NOT simply chaos, which is easily described by its average.
- Similarly, power law's are not well defined by their average. They should be indicative of a system forced to mirror another complex process.
- Also seems opposite to earlier 'no apriori' angle, where power law is the initial assumption. However, power-law-like behavior (assoc even to multiplicative noise) should be observed on parity-like problems (result, not init assumption).
- Power law may not be network trained on parity problems, but network trained across all problems (exp weighted for rare events such as parity). Does power law imply that networks of this type *become* like the parity problem? Would make sense, since should not be any shortcut to describing them.

Misc modified ideas

- nodes w/ higher info are more stable, ie mutate less (do not have higher pr of forming new edges)
- starting point for new node should be same as starting point for whole net
- division as splitting a node into smaller subsets (fracturing, top down)
    - when does division occur? opposite of merging

- from Edges as Pointers:
    - form new edge if represn of base signal reqs > mem than an edge?
    - "cost of ptrs increase as net grow, but due to #ordered combos, the cost increases very slowly"
    - "each node has some sort of internal represnetation of others, whose composition may differ from others' internal representation"

Earlier Proposal

Want to move twds algo that has the minimum run time over X,Y, which represents all possible instances of x, y.

X,Y ~ x,y
x ~ {{0,1}^{m}}^{n}, y ~ {0,1}^{n}
Assume instances $(x_i, y_i) \in (X, Y)$ are drawn IID (input instances rd shuffled) (indpd along n)
However, inputs of any $(x_{ij}, y_i) \in (x_i, y_i)$ are NOT nec IID (x's and y's may have depds and vary) (dpds along m)

Normally assume IID and use Bayes Rule: (unclear what optimizing over, i guess determine subset of $x'_j s$ to include)
$\max << P(x_j = x_{ij} | y = y_i) >_{(x_{ij}, y_i)} >_{(x_i, y_i)}$ -> loglikelihood estimation

$\max \prod_i < P(x_j = x_{ij} | y = y_i) >_{(x_{ij}, y_i)}$

if $x'_j s$ were independent could use
$\max \prod_i \prod_j P(x_j = x_{ij} | y = y_i)$

Can see why rare events, ie certain $(x,y) \in (X,Y)$ drastically reduce likelihood (hence rare derr). (conflating scales...)
But this assumes IID between x's, ie no multivariate depds. Instead let dependency graph G represent some ordering of dependencies between x's, without cycles [how to write this mathematically? same for below]

Then
$\max < P(x = x_i \cap G | y = y_i) >_{(x,y) \in (X,Y)}$

$\max << P(x = x_i j \cap G | y = y_i) >_{(x_{ij}, y_i)} >_{(x_i, y_i)}$

Easy typical runs vs hard rare runs...how to explain mathematically...?

Very ordered as $y = x_i$. Verry disordered as $y = avg(x_i)$. Complex as inbetween.

Parity as prototypical rare event. Most resistant to description via the mean. mi persp: all instances of x_i and organize are evenly split by their corresp y. MA persp: plot density of instances of X and evenly split by corresp y. *Parity is minimally dependent on individual x_i (indepd indivly), while maximially sensitive to any change in x_i.*

Info as relevant info far from mean (which is already the a priori assumption). Already has log scaling (as response to exp weights?).

No closed form solution (v hand wavy atm). And don't want to check all combos.

Want max syngetic info captured, while min redundant info. These two sides should follow the same algorithm:

- add new (syn) or remove existing (rdt) edges by comparing against others in a node
- divide nodes pr'ly (usually in half): if syn separate them, if rdt kill the weaker one