The Alan Turing Institute

Counting Small Patterns and Testing for Independence

Chaim Even-Zohar

Joint work with Calvin Leng

Counting Patterns

#σ(π) = how many occurrences of **σ** in **π**

Counting Patterns

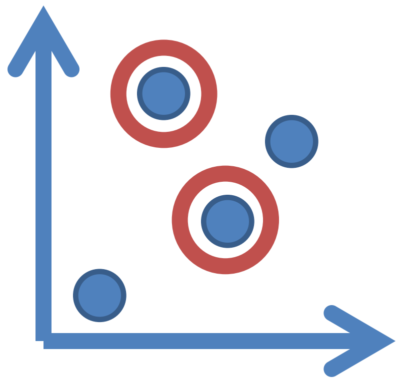
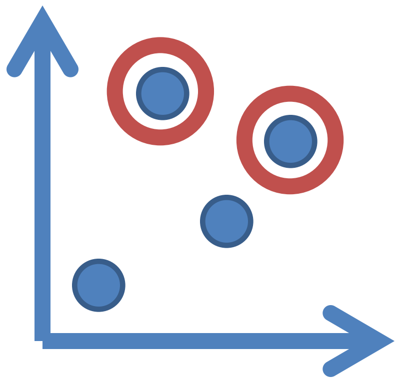
$\#\sigma(\pi)$ = how many occurrences of σ in π

$\#21(1423) =$

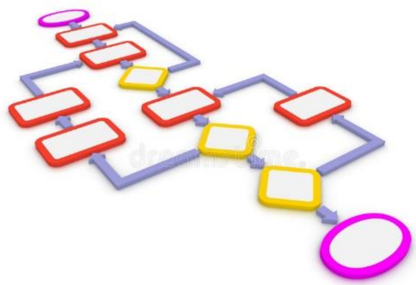
Counting Patterns

#σ(π) = how many occurrences of **σ** in **π**

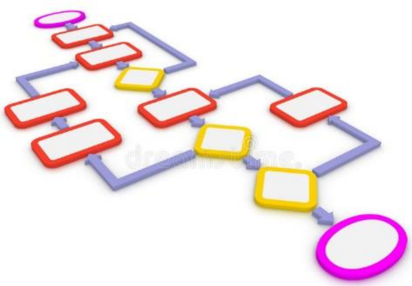
#21(1423) = 2

Counting Patterns

σ, π →  → $\#\sigma(\pi)$

Counting Patterns

σ, π →  → $\#\sigma(\pi)$

#P-complete for large $n=|\pi|$, $k=|\sigma|$

[Bose Buss Lubiw 1998]

ETH => not $n^{o(k / \log k)}$

[Berendsohn Kozma Marx 2019]

Important Special Cases

#1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

Important Special Cases

#1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

Can tell quasirandomness / independence

[Hoeffding 1948] [Blum Kiefer Rosenblatt 1961] [Yanagimoto 1970] [Bergsma Dassios 2010] [Král' Pikhurko 2013] [Chan Král' Noel Pehova Sharifzade Volec 2020]

Important Special Cases

#1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

Can tell quasirandomness / independence

[Hoeffding 1948] [Blum Kiefer Rosenblatt 1961] [Yanagimoto 1970] [Bergsma Dassios 2010] [Kráľ Pikhurko 2013] [Chan Kráľ Noel Pehova Sharifzade Volec 2020]

So far best known: $O(n^2)$ time

[Weihs Drton Leung 2016] [Heller Heller 2016]

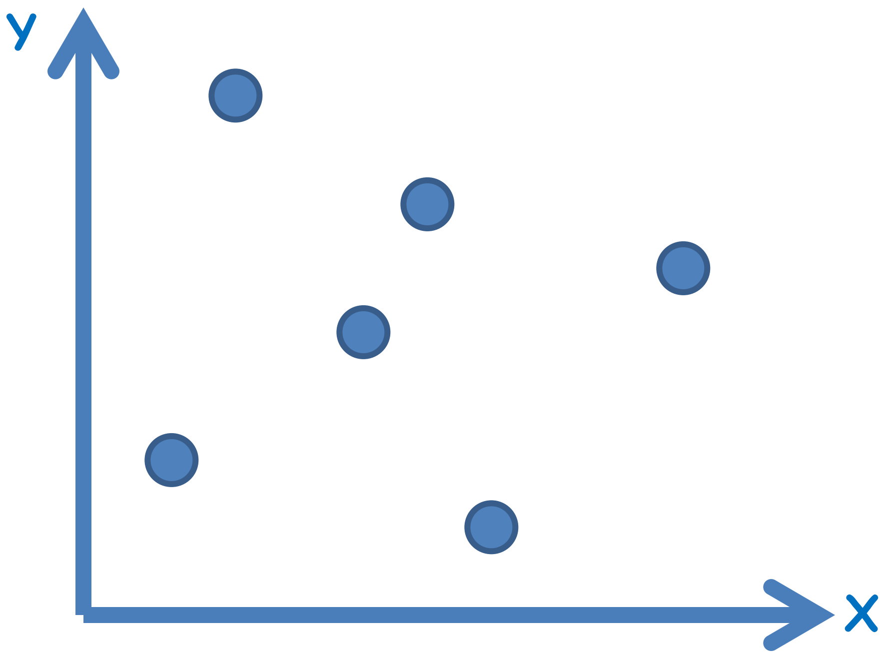
Bivariate Data

X, Y non-atomic random variables

Bivariate Data

X, Y non-atomic random variables

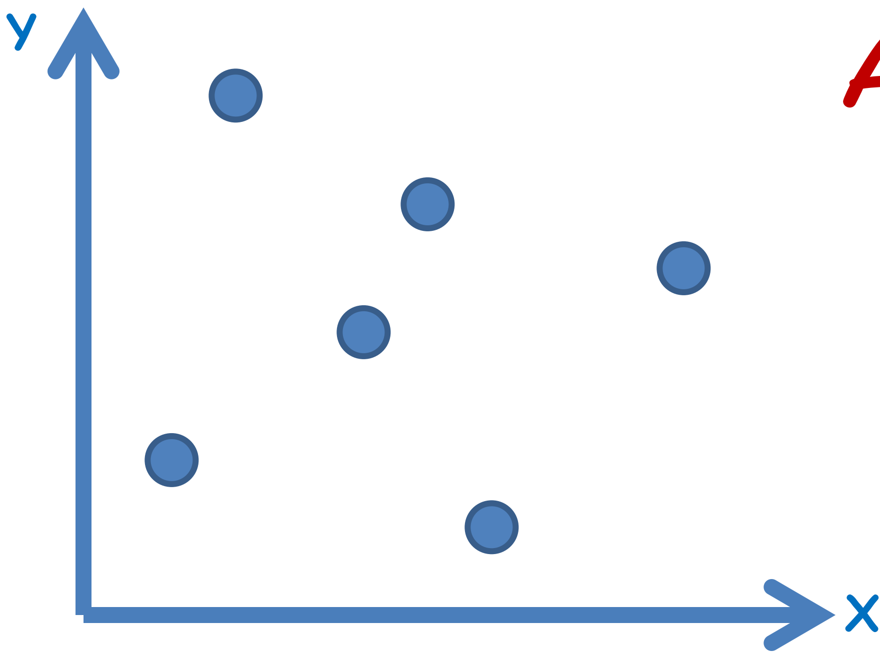
$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ iid samples



Bivariate Data

X, Y non-atomic random variables

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ iid samples

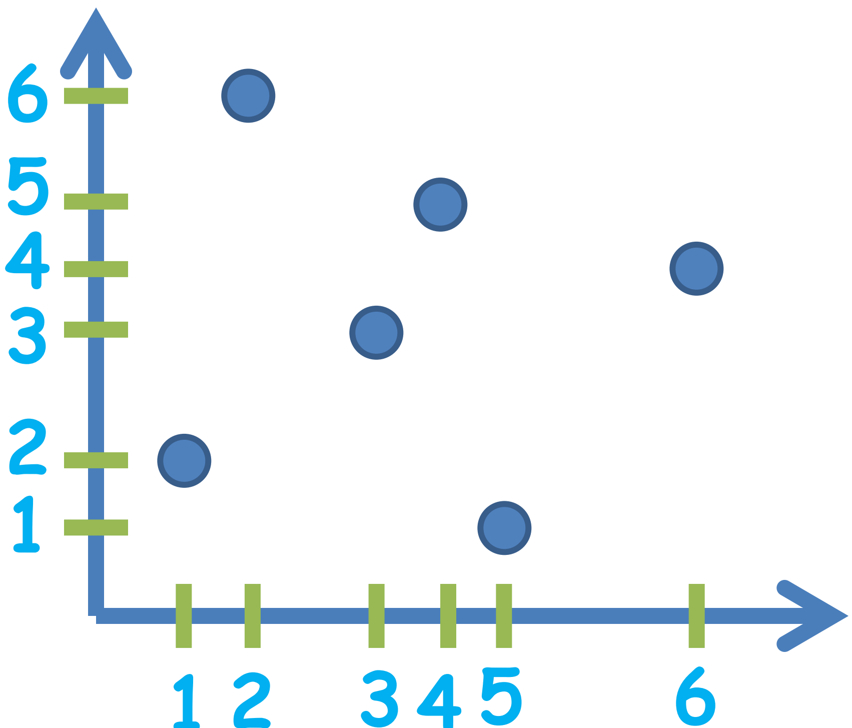


Are X and Y independent?

Bivariate Data

X, Y non-atomic random variables

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ iid samples



Are X and Y independent?

$\pi(\text{ rank } X_i \text{ }) = \text{ rank } Y_i$

$\pi = 2\ 6\ 3\ 5\ 1\ 4$

Results

τ_n = #1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

Theorem E, Leng 2019 (SODA 2021)

τ_n is computable in **O(n log n)** time

Results

τ_n = #1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

Theorem E, Leng 2019 (SODA 2021)

τ_n is computable in **O(n log n)** time

=> Consistent and Near-linear Hoeffding-type Independence Tests

Results

τ_n = #1234 + #1243 + #2134 + #2143

+ #3412 + #3421 + #4312 + #4321

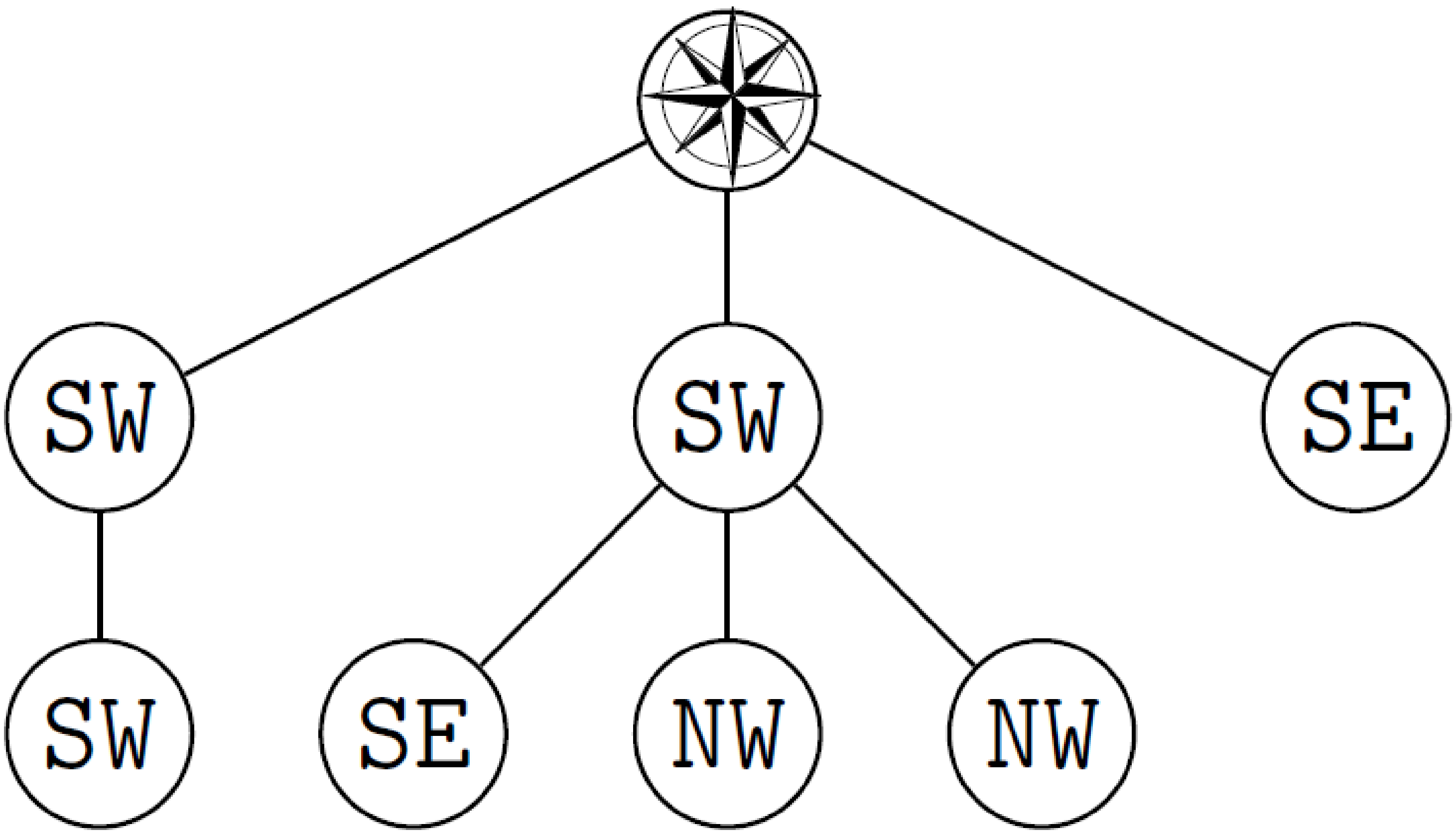
Theorem E, Leng 2019 (SODA 2021)

τ_n is computable in **O(n log n)** time

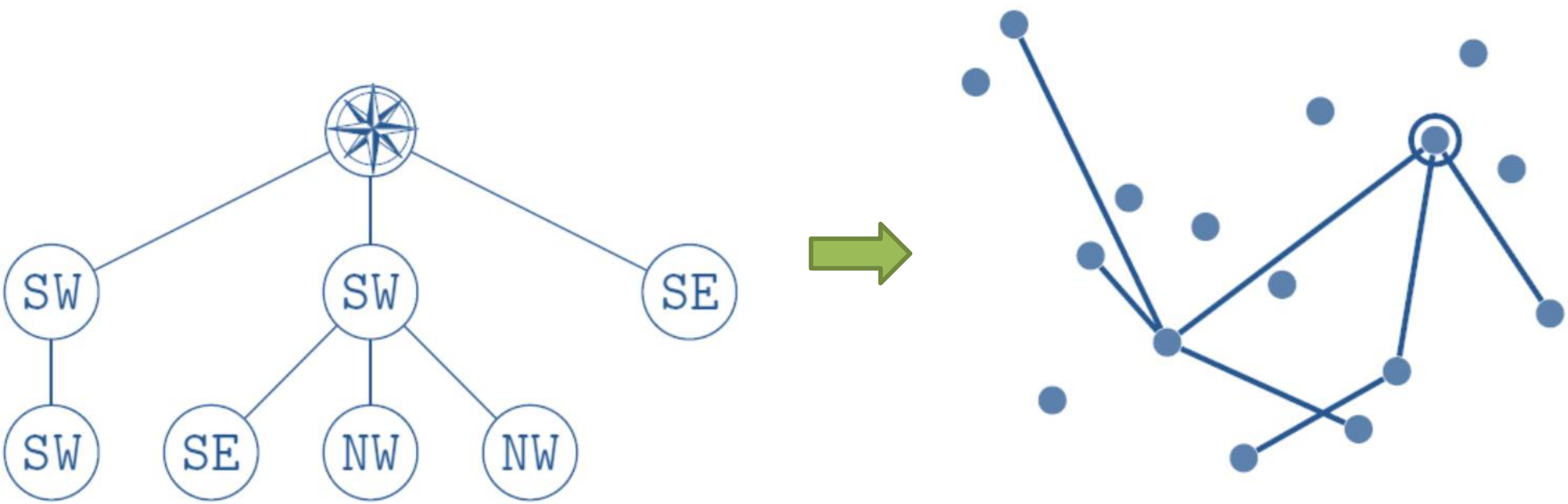
Moreover, in O(n log n) time:

- Each one of these eight pattern counts
- A **23**-dim subspace of 4-pattern combinations

Corner Trees

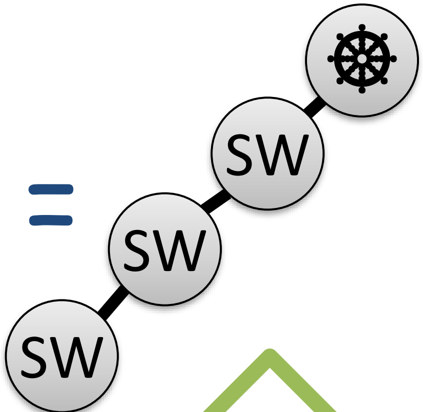


Corner Trees

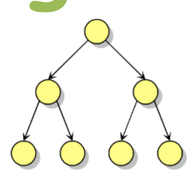


Corner Tree Formulas

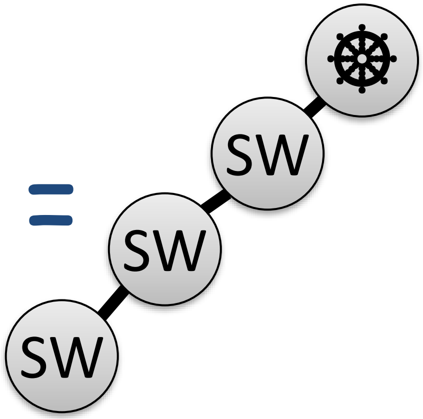
1234 =

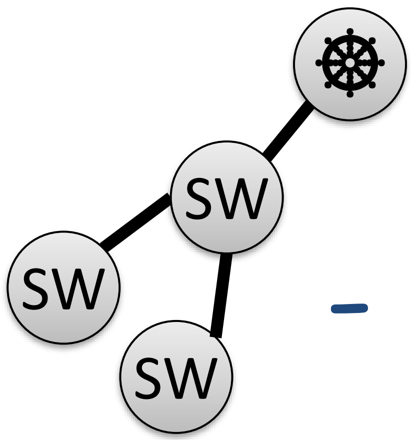


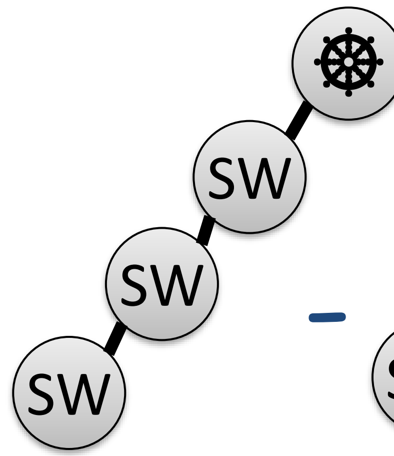
Dynamic Programming Algorithm

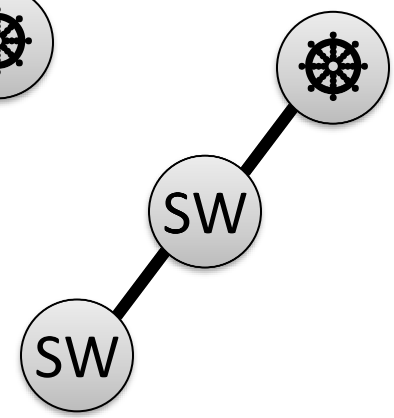


Corner Tree Formulas

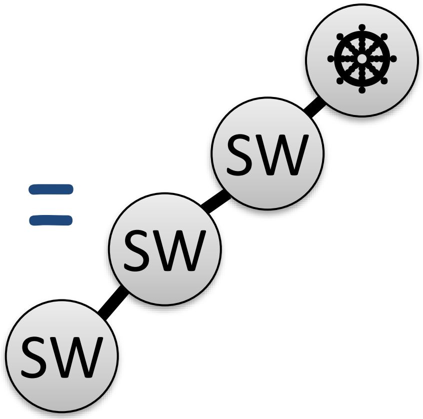
1234 = 

2134 = $\frac{1}{2}$ 

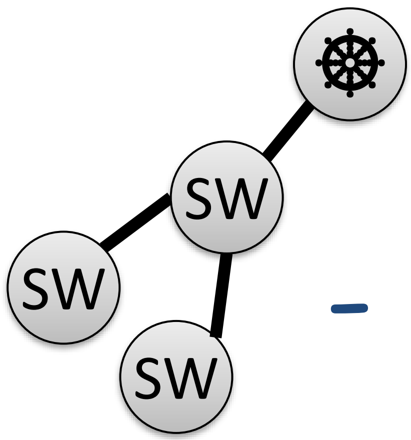
$- \frac{1}{2}$ 

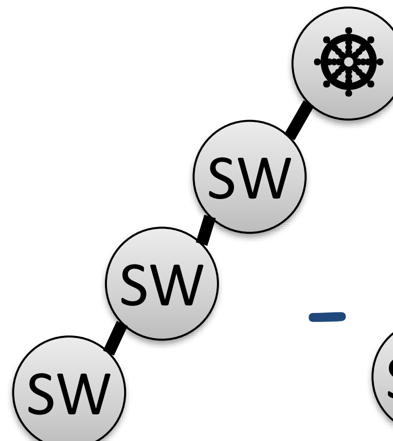
$-$ 

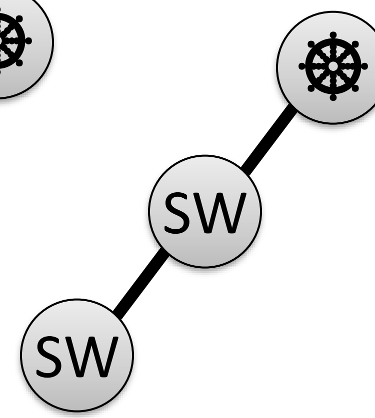
Corner Tree Formulas

1234 = 

4321 similar

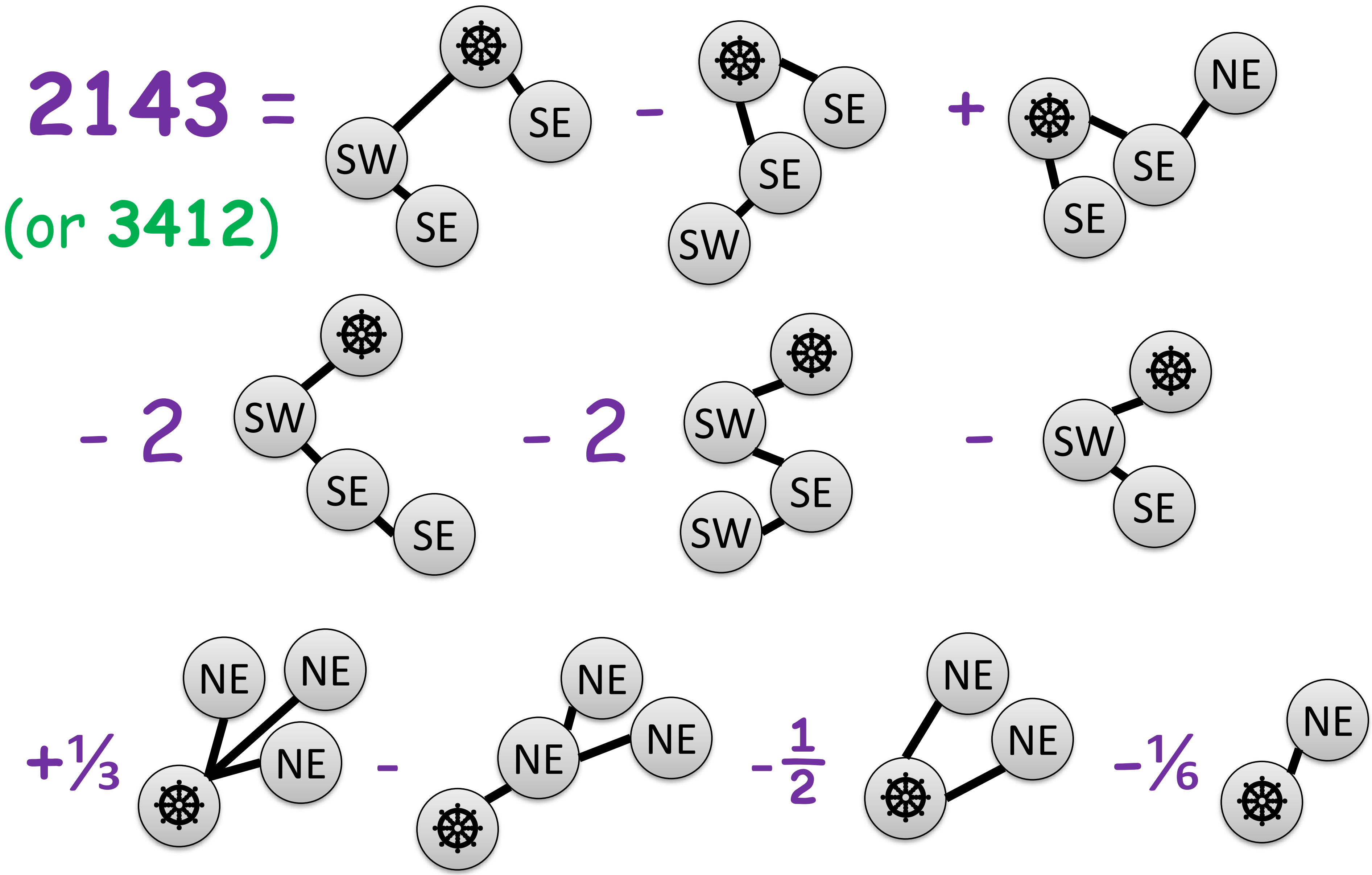
2134 = $\frac{1}{2}$ 

$-\ \frac{1}{2}$ 

$-$ 

1243, 3421, 4312 similar

Corner Tree Formulas



THANKS