数据流分析

理论基础

钮鑫涛

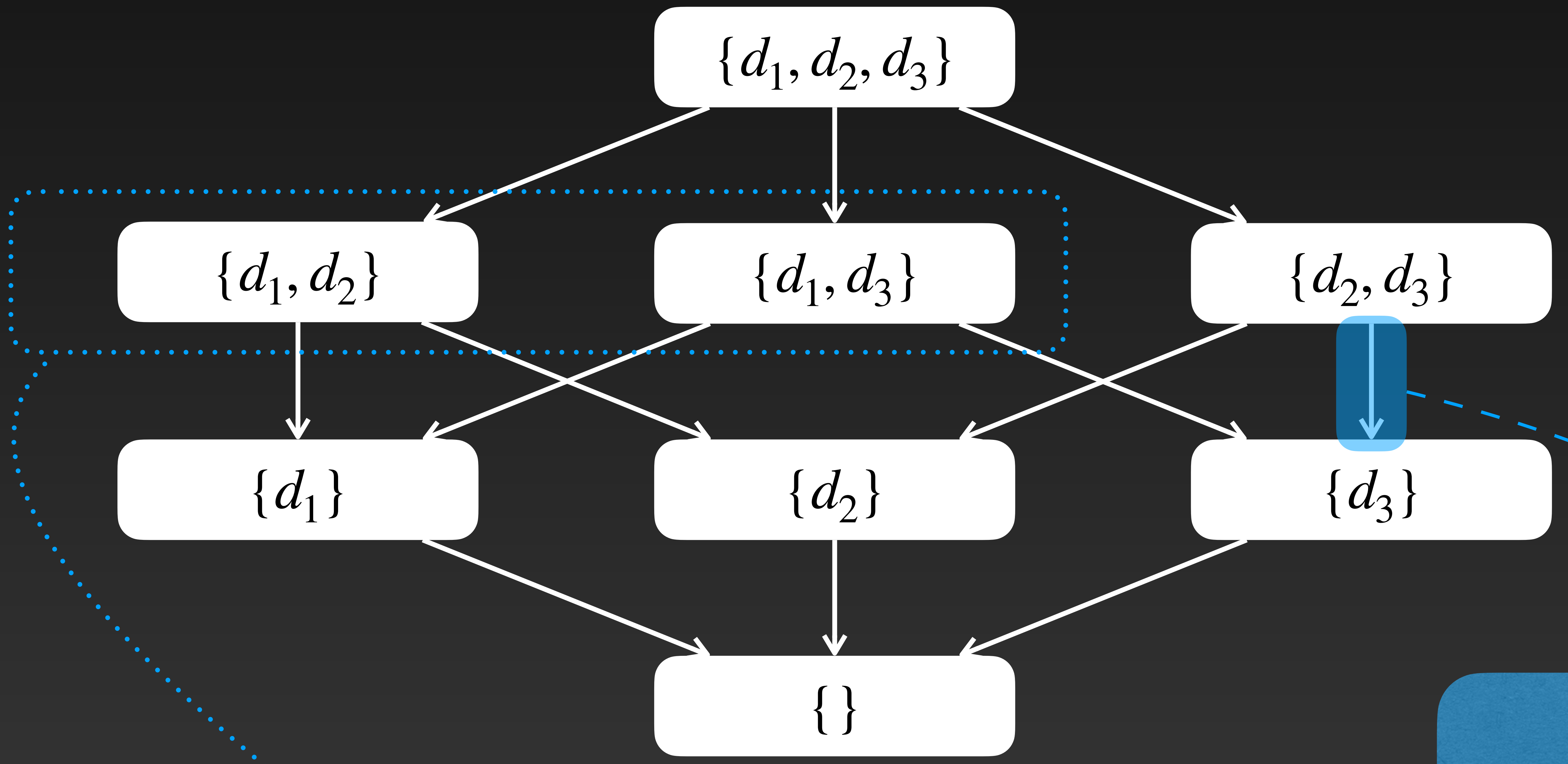
一些关于“序(Order)”的知识

偏序集（Partial Order Set, POSET）

- 一个偏序集$(P, \preceq)$是一个集合P和一个二元关系$\preceq$, 对所有的$x, y, z \in P$有：
    - 传递性（Transitive）： 如果 $x \preceq y$并且$y \preceq z$, 那么$x \preceq z$
    - 反对称性 (Anti-symmetric): 如果 $x \preceq y$并且$y \preceq x$, 那么$x = y$
    - 自反性（Reflexive）: $x \preceq x$

偏序集例子

- 令$V = \{x \mid x \subseteq \{d_1, d_2, d_3\}\}$, $\preceq \, = \, \subseteq$, 我们有如下图的偏序集（哈斯图表示）



有一些元素不符合二元关系，因此不是全序

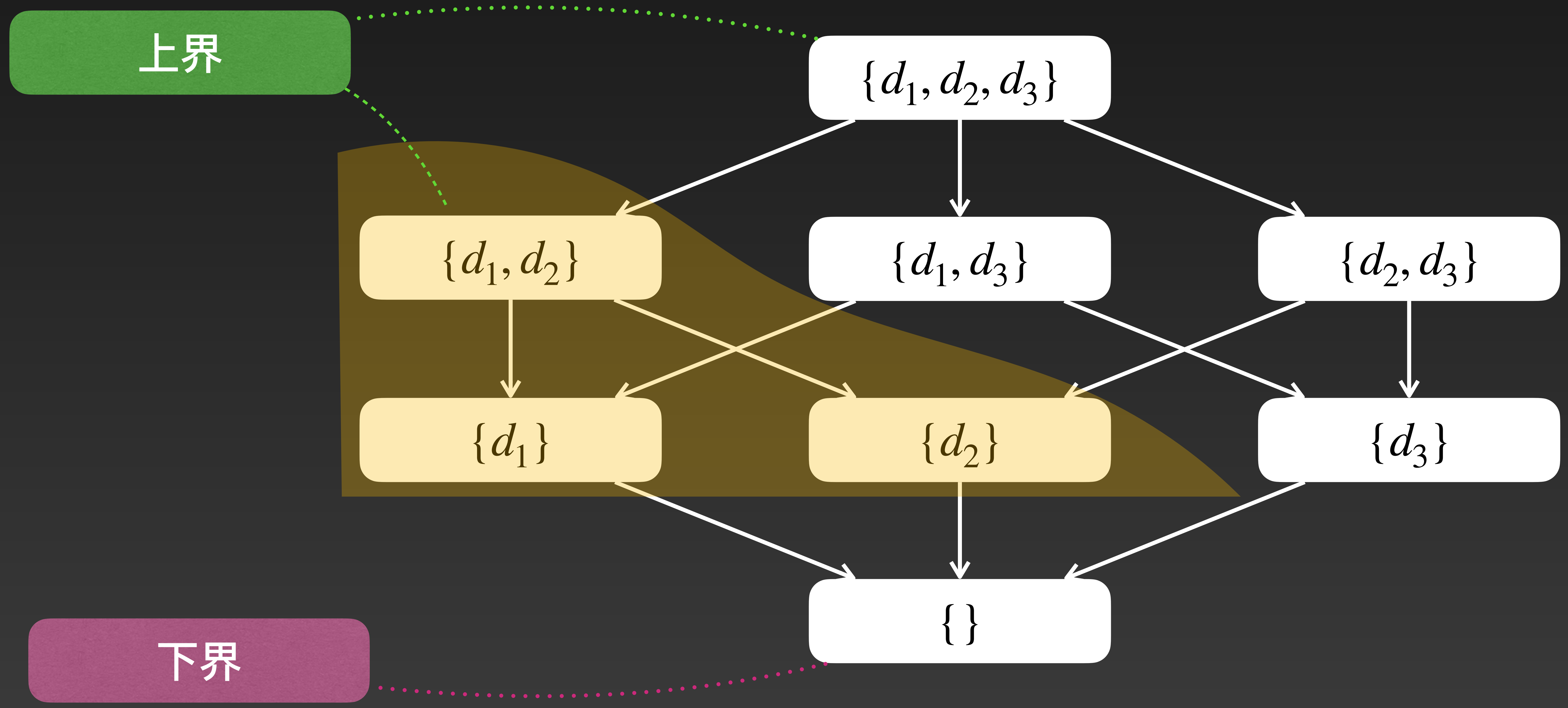
这里边表达了“覆盖”（Covers）关系：

$x \lessdot y \triangleq (x \prec y)$ and $\neg(\exists z \,.\, x \prec z$ and $z \prec y)$

即$\lessdot$只表达“直接”的二元关系，并且没有自指的关系

上界（upper bound）和下界（lower bound）

- 给定一个偏序集$(V, \preceq)$，给定V的一个子集$S \subseteq V$, 我们有$u \in V$是S的一个上界，当且仅当$\forall x \in S, x \preceq u$. 注意：$u$不必在$S$内！同样，我们有$l \in V$是$S$的一个下界，当且仅当$\forall x \in S, l \preceq x$.

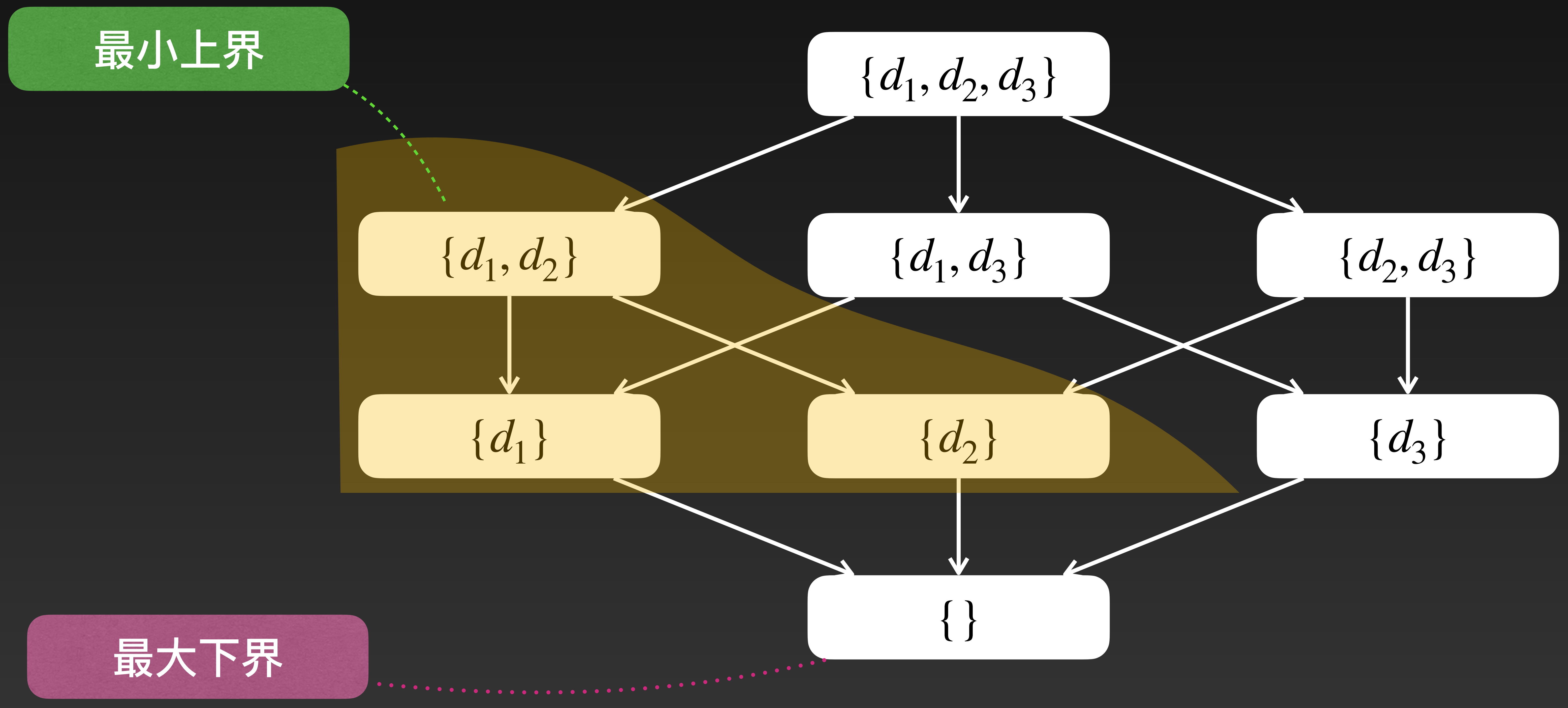


最小上界（least upper bound）和最大下界（greatest lower bound）

- 一个偏序集S的上界x被称为S的最小上界（least upper bound，lub）或紧上界（tight upper bound）或者上确界（supremum）当且仅当任何一个S的上界都比该上界还大（∀upper bound y of S, $x \preceq y$）
  - 可以表示为lub(S)，$\vee S$ 或者$\sup S$
- 一个偏序集S的下界x被称为S的最大下界（greatest lower bound，glb）或紧下界（tight lower bound）或者下确界（infimum）当且仅当任何一个S的下界比该下界还小（∀lower bound y of S, $y \preceq x$）
  - 可以表示为glb(S)，$\wedge S$ 或者 $\inf S$

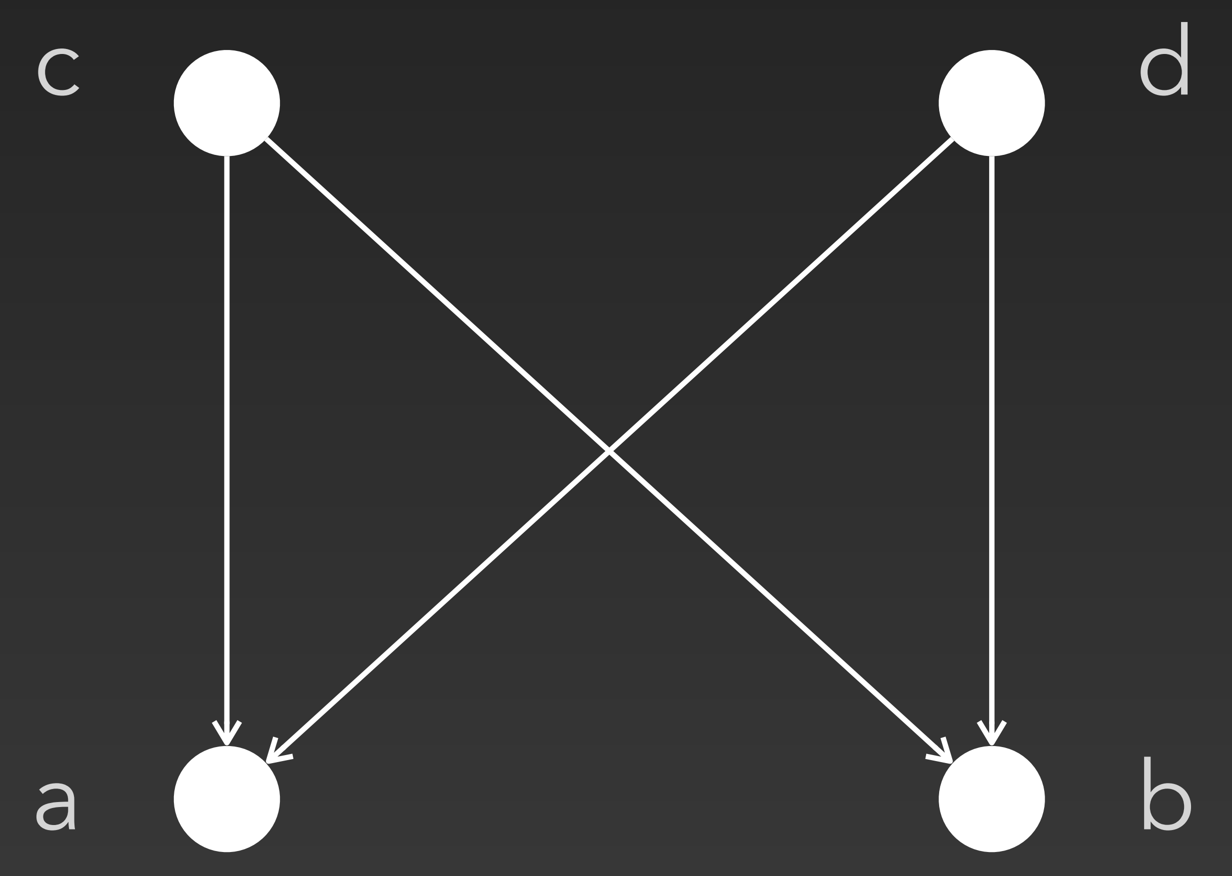
上确界性在数学中很重要，是实数定义的基本公理（其保障了实数没有“gap”，也被称为Dedekind completeness）

最小上界（least upper bound）和最大下界（greatest lower bound）



最小上界（least upper bound）和最大下界（greatest lower bound）

- 如果一个偏序集S存在一个最小上界lub(S)，并且该元素属于S，即lub(S) $\in S$，则称lub(S)为其最大元（greatest element），记为T（top element）
  - 显然lub(S)如果存在即唯一！否则不符合最小上界的定义
  - 对于一个偏序集而言，lub(S)不一定存在，如下面的例子



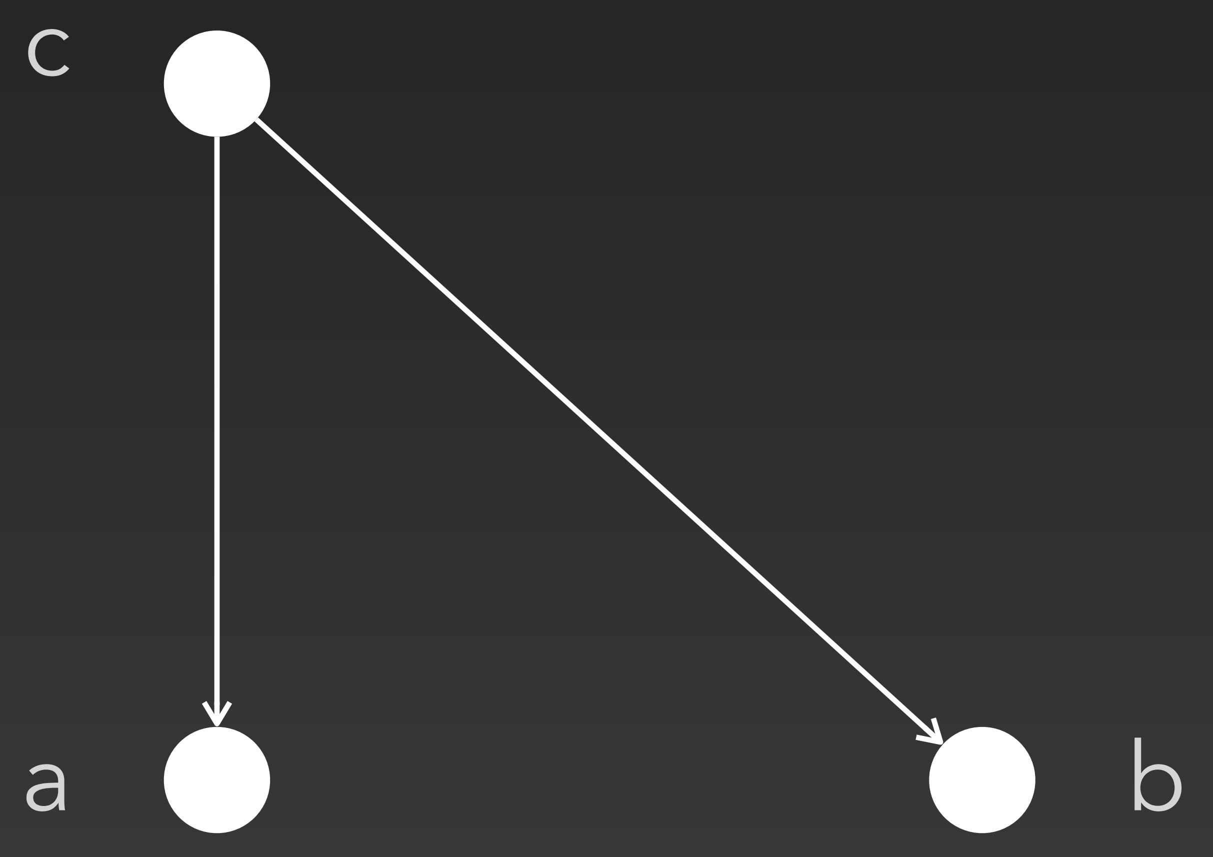
$$a \preceq c, a \preceq d, b \preceq c, b \preceq d$$

对于subset {c, d}而言， 其甚至都没有一个 upper bound!
也就是不用说least upper bound了

对于subset {a, b}而言，其有两个upper bounds, 分别为c和d，但是没有least upper bound，也就不存在最大元了

最小上界（least upper bound）和最大下界（greatest lower bound）

- 如果一个偏序集S存在一个最小上界lub(S)，并且该元素属于S，即lub(S) $\in S$，则称lub(S)为其最大元（greatest element），记为T（top element）
  - 显然lub(S)如果存在即唯一！否则不符合最小上界的定义
  - 有最小上界，也不一定有最大元，如下面的例子



$$a \preceq c, b \preceq c$$

对于subset {a, b}而言，其有一个upper bound, 为c，即c是最小上界，但是c不在{a, b}中，因此{a, b}没有最大元

另一个例子：对于实数子域 [0, 1) 而言，其有最小上界，但没有最大元

最小上界（least upper bound）和最大下界（greatest lower bound）

- 类似地（对偶地），如果一个偏序集S存在一个最大下界glb(S)，并且该元素属于S，即lub(S) $\in S$,则称glb(S)为其最小元（least element），记为⊥（bottom element）
  - 同样，glb(S)如果存在即唯一！否则不符合最大下界的定义
  - 对于一个偏序集而言，glb(S)不一定存在
  - 有最大下界，也不一定有最小元

极大元和极小元

- 极大元 (Maximal element)
  - 对于一个偏序集S而言，一个元素$m \in S$是极大元(maximal element) 当且仅当S中不存在比其"大"的，即$\forall s \in S$, 如果$m \preceq s$，那么必有$s \preceq m$, 极大元集合记为$\max(S)$
    - 注：m不必比任何元素都大，这是和最大元（greatest element）不同的地方！因此其也不必唯一
- 极小元 (Minimal element)
  - 对于一个偏序集S而言，一个元素$m \in S$是极小元(minimal element) 当且仅当S中不存在比其"小"的，即$\forall s \in S$, 如果$s \preceq m$，那么必有$m \preceq s$, 极小元集合记为$\min(S)$
    - 同样地，m不必比任何元素都小，这是和最小元（least element）不同的地方. 其也不必唯一

Theorem. Any finite poset has at least a maximal/minimal element. (can by proved by induction on the size of poset)

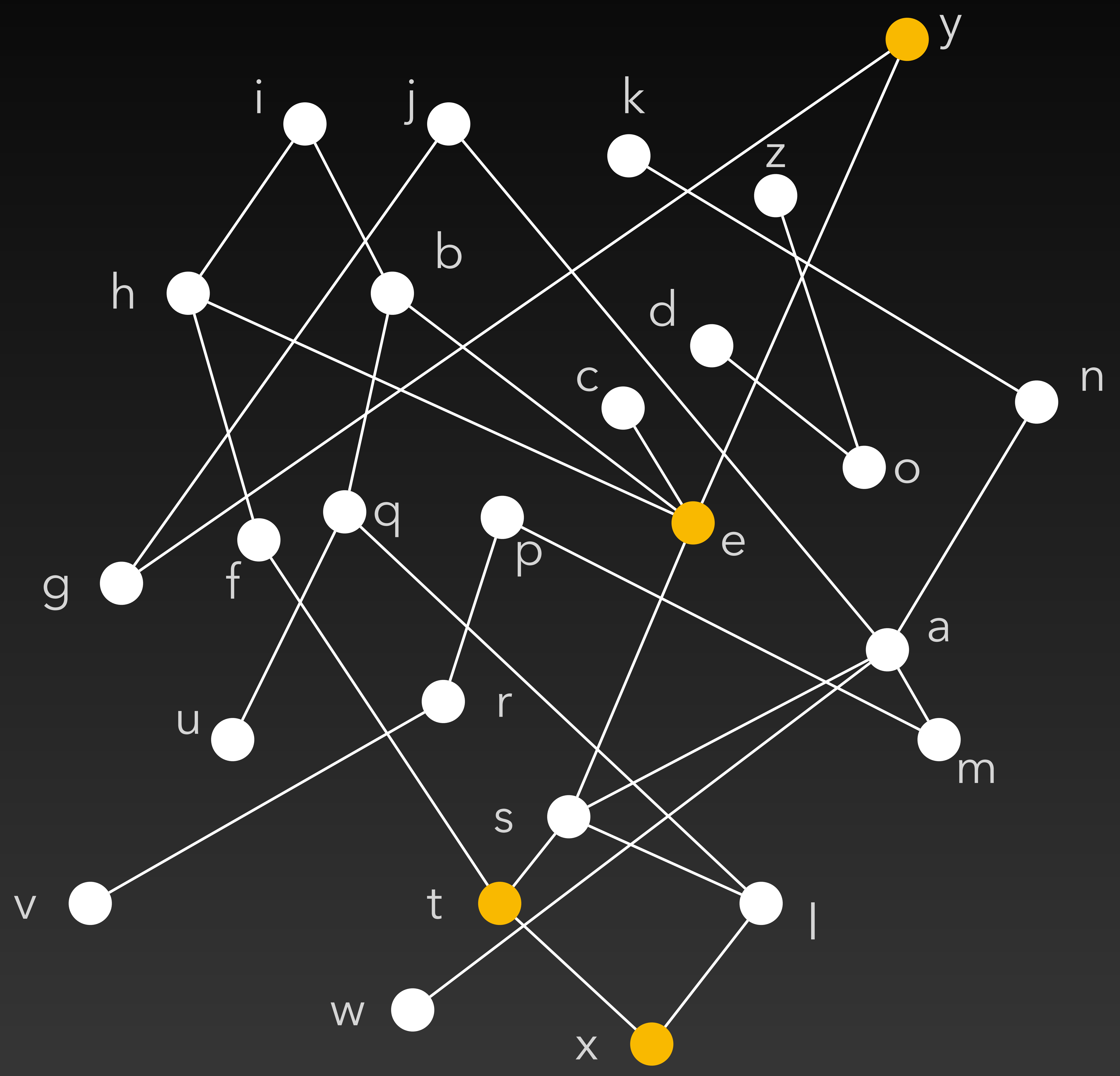
链（Chain）

- 链（Chain）
  - 一个偏序集被称为链当且仅当其中的任何连个元素都是可以比较的（comparable），即其是一个全序集（total ordered set）或线性序集（linearly ordered set）
    - 所谓的可比较即可以用≤关系描述（$x \preceq y$, 或者$y \preceq x$）,因此一条链就是一个类似$x_1 \preceq x_2 \preceq x_3 \ldots \preceq x_s$的序列
  - 一个偏序集P的非空子集C是P的一个链，当$(C, \preceq)$本身是一个链，P所包含的所有链集记为Chn(P)
  - 一个偏序集P的链C是最大的（maximal），当且仅当不存在P的其他链C',满足$C \subsetneq C'$
  - 一个偏序集的高度（height）就是这个偏序集中最长的链（多个maximal chain里最长的那个）所含有的元素个数（即基，cardinality）

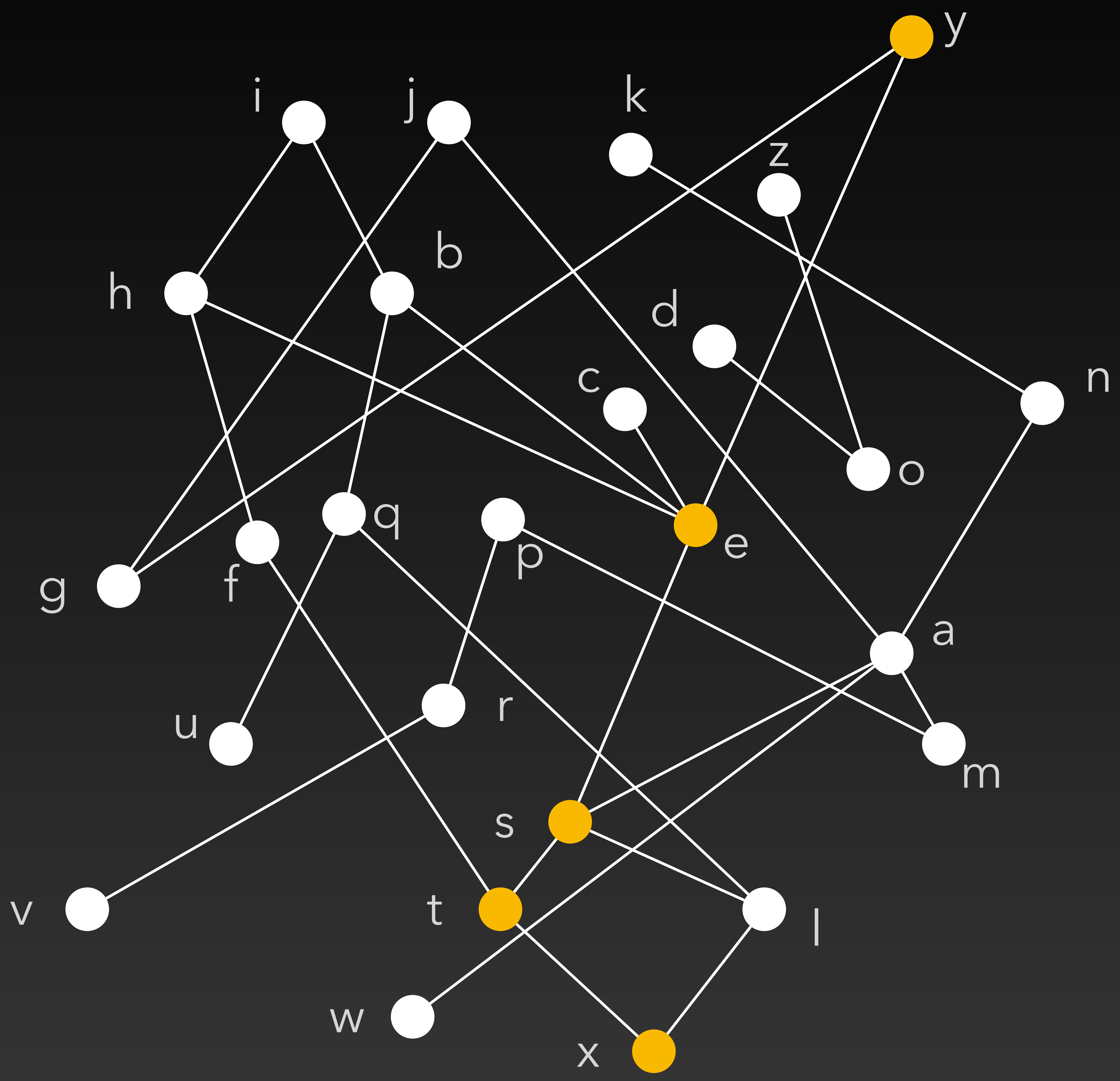
Zorn's Lemma: 给定一个偏序集P，其所含的每一条链都有一个上界，那么P包含至少一个极大元(maximal element)

（Zorn's Lemma 给定了一个偏序集含有极大元的充分条件，其不限制P是否有限，Zorn's Lemma不可证，其等价于选择公理Axiom of Choice）

链（Chain）



Chain



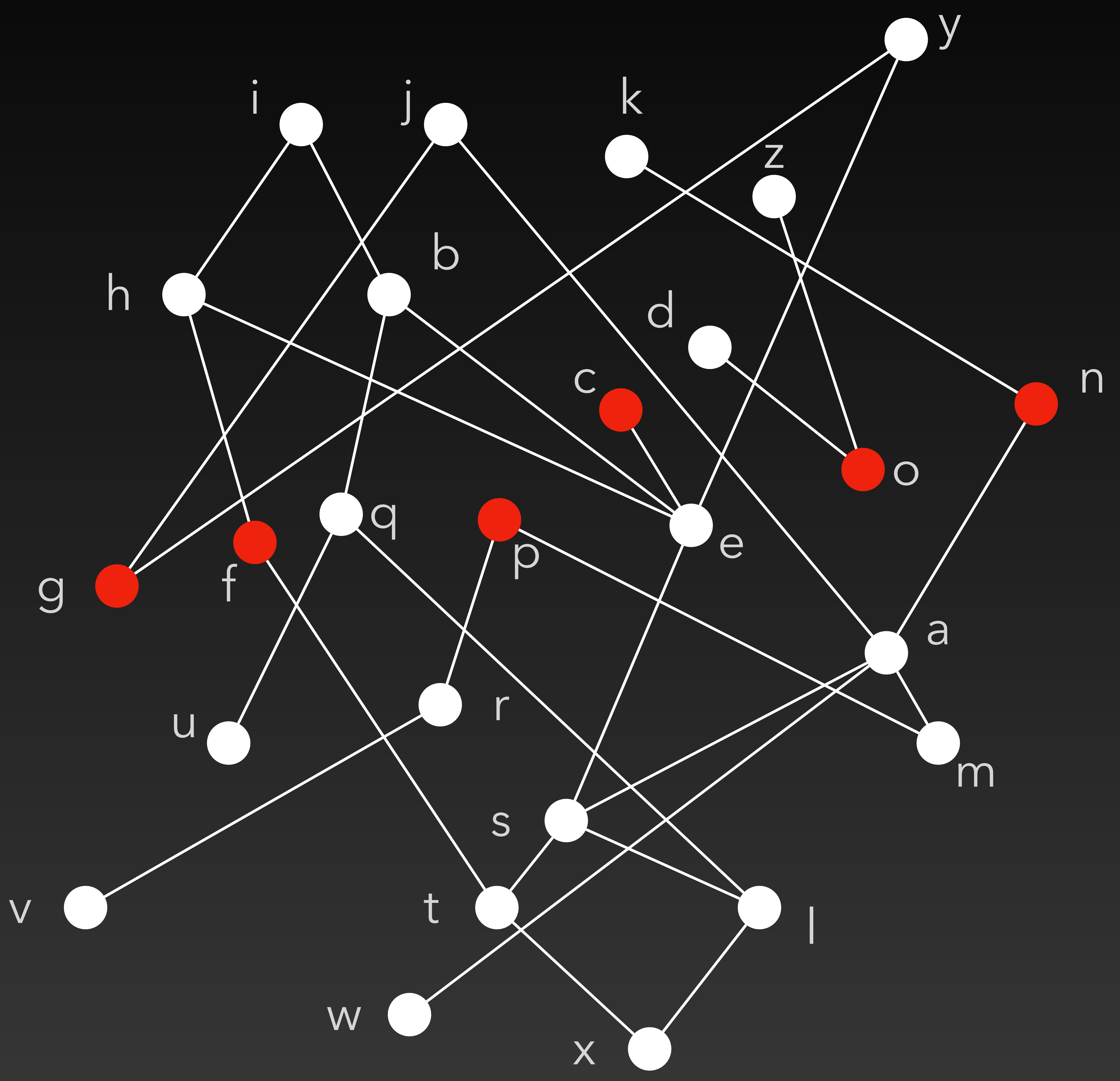
Maximal Chain

反链（Anti-Chain）

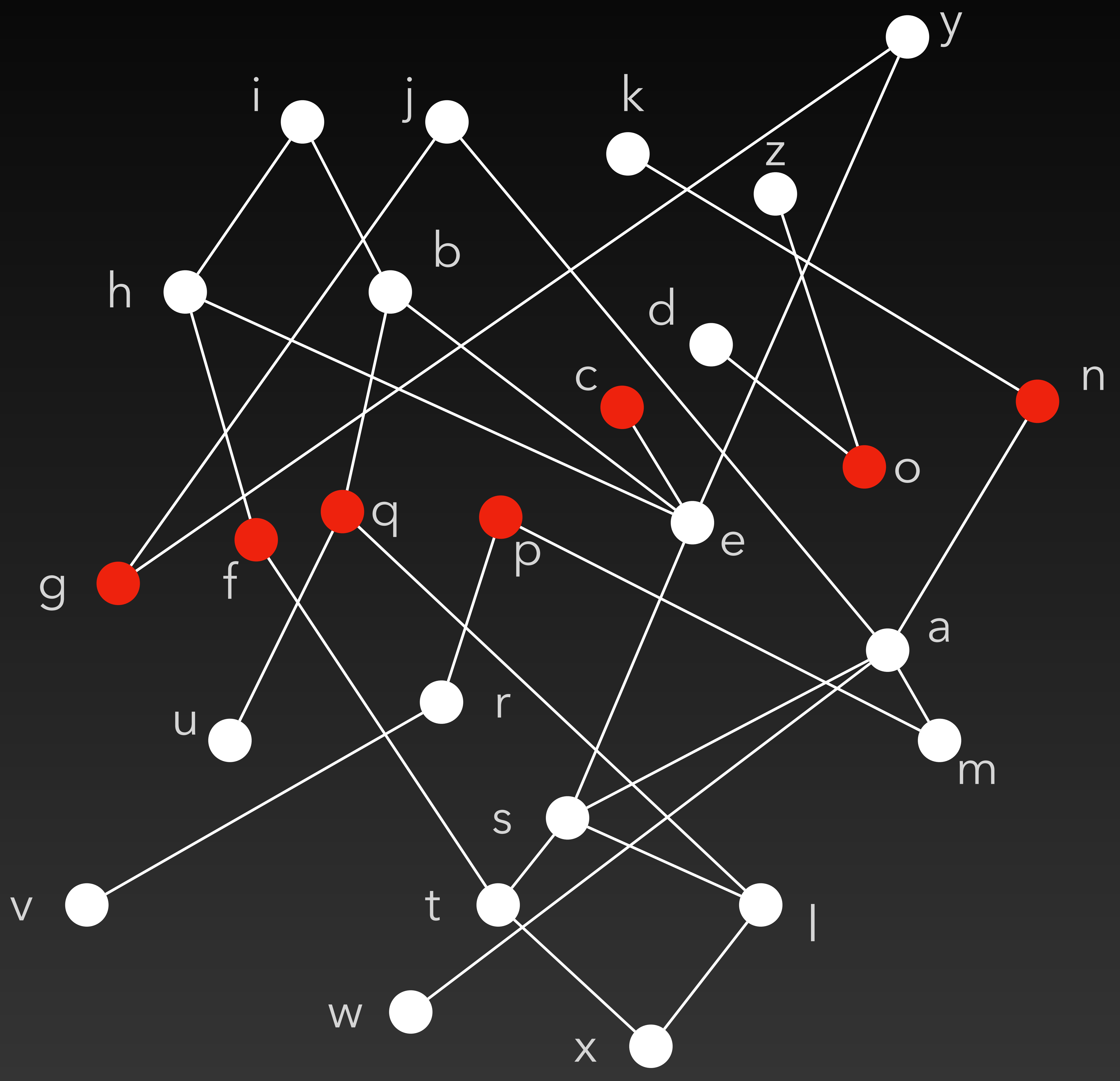
- 反链（Anti-Chain）
  - 一个偏序集P的子集A，被称为P的反链（Anti-Chain，也称为Sperner族）当且仅当对于A中的任意两个元素，他们都不可比较.
    - 一个P的所有反链形成一个P的反链集，记为Anti(P)
    - 一个偏序集P的反链C是最大的（maximal），当且仅当不存在P的其他反链C',满足$C \subsetneq C'$
  - 一个偏序集的宽度（width）就是这个偏序集中最长的反链的基（元素个数）

对于一个有限的Poset而言，其一个链和一个反链最多有一个共同的元素！

反链（Anti-Chain）

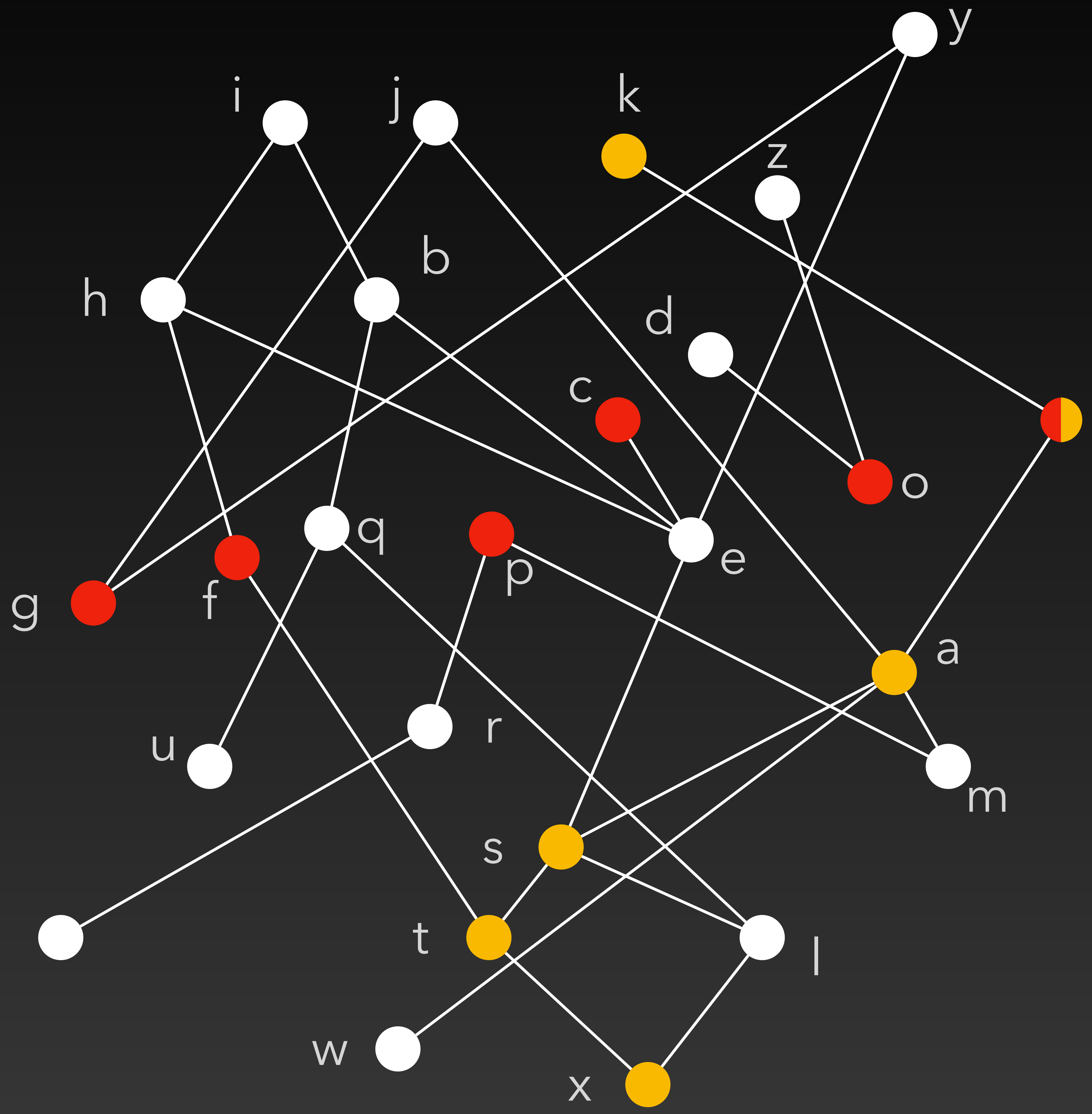


Anti-Chain



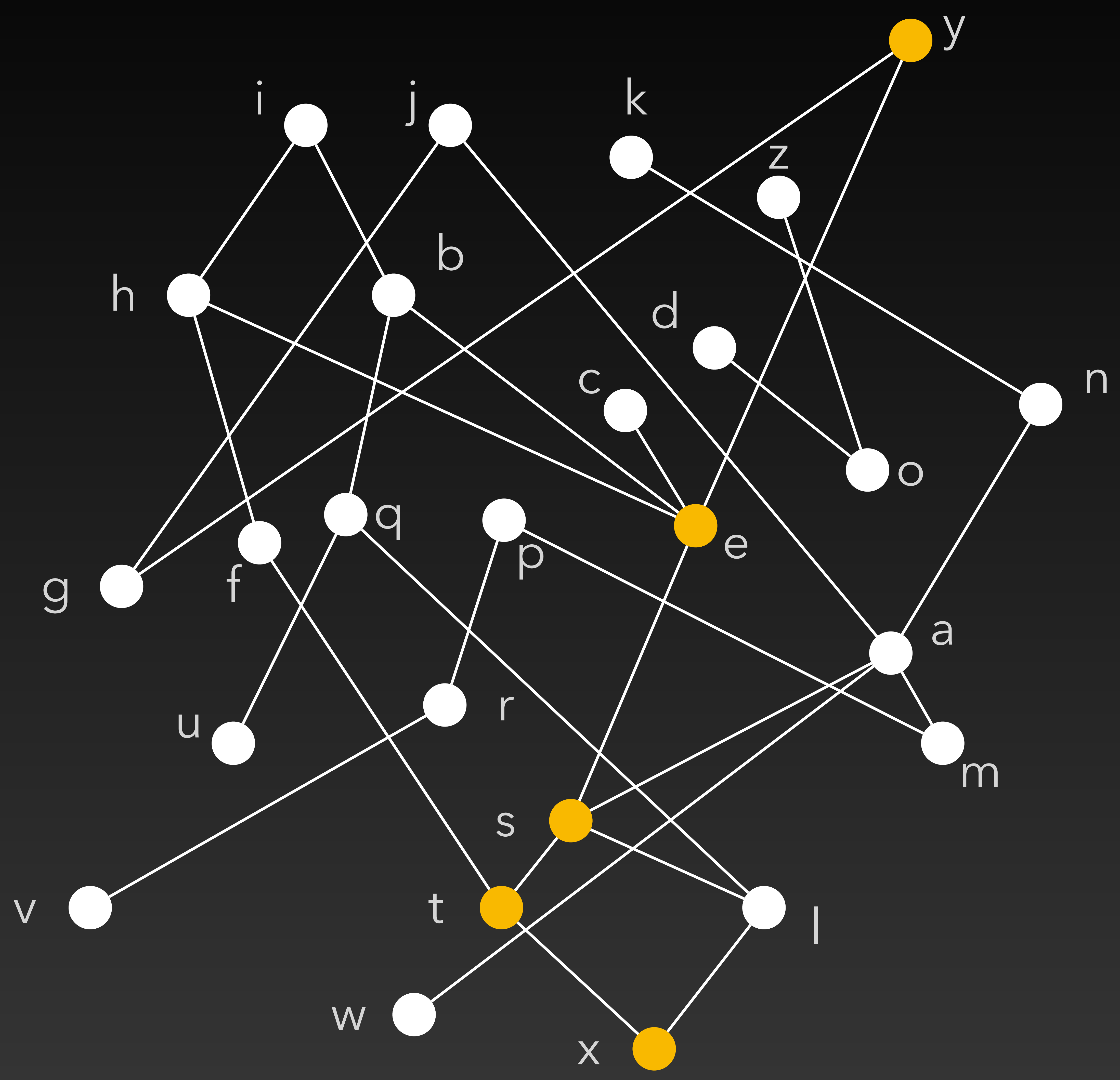
Maximal Anti-Chain

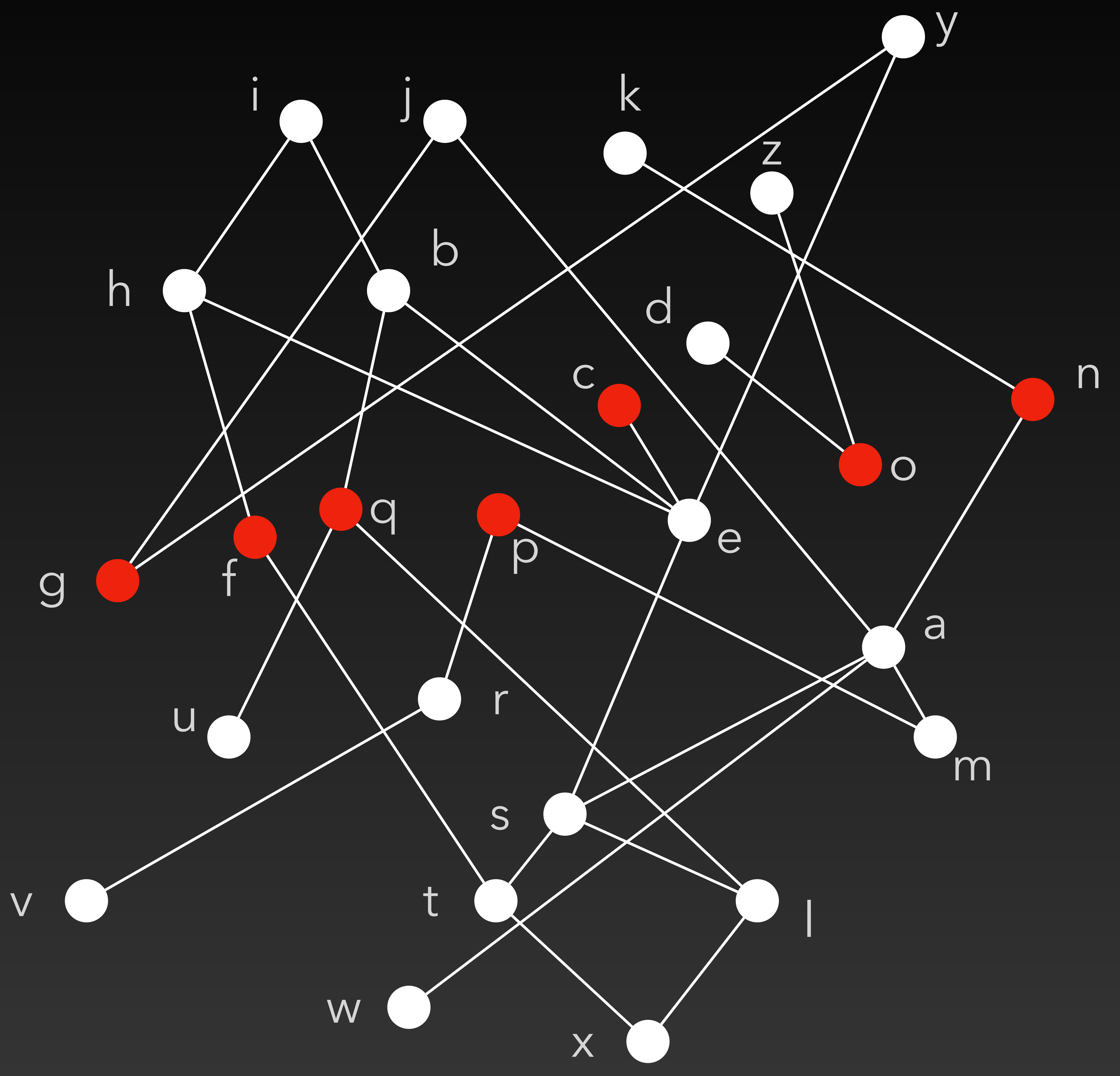
链和反链



一个链和一个反链最多有一个共同的元素！

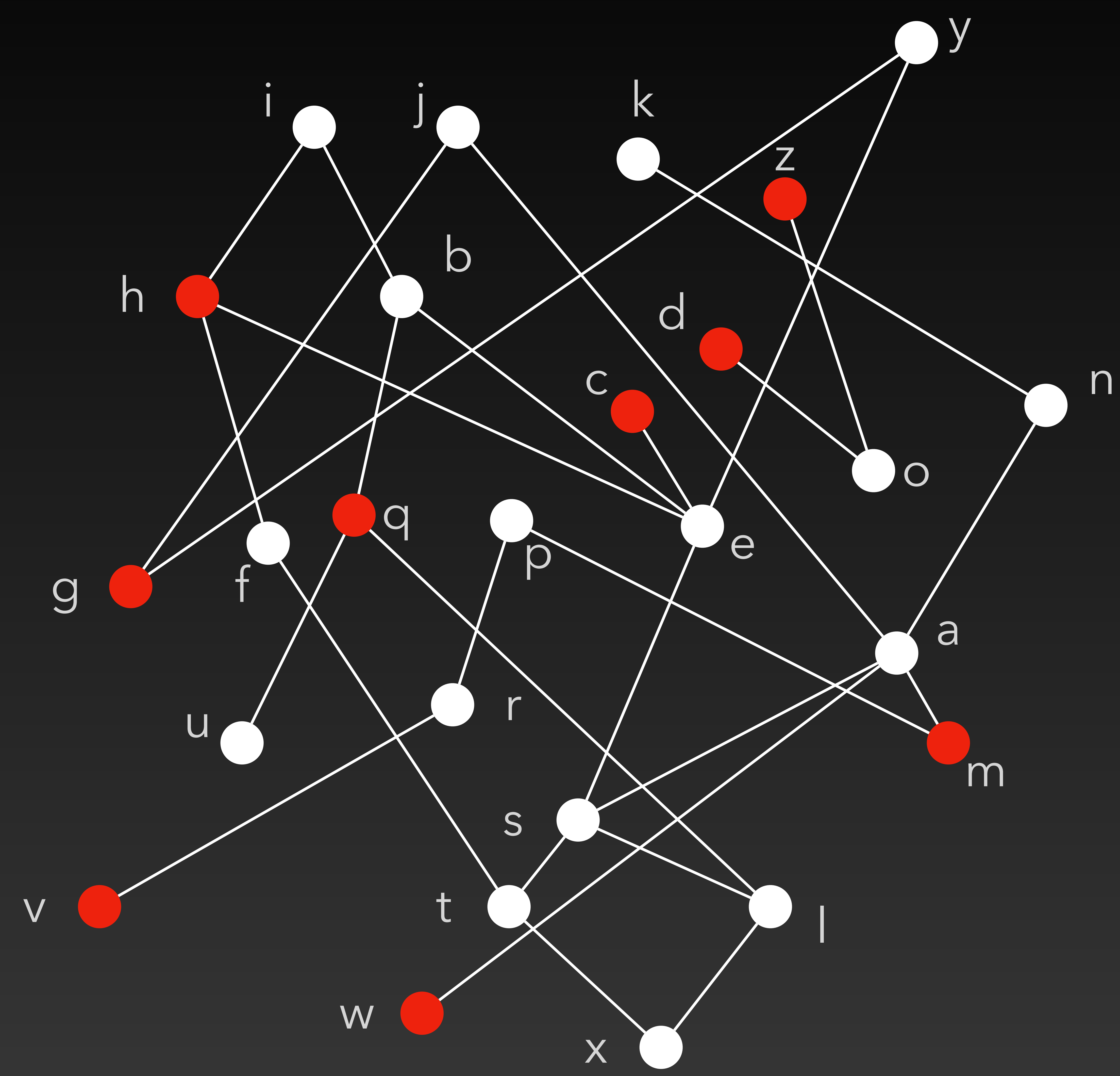
链和反链





问题： 一个POSET的高度和宽度如何得到？一个简单的回答是给出一个maximal chain/anti-chain的样例, 那么高度/宽度至少大于等于这个样例的元素个数，但是如何能确定这个样例就是最长链/反链呢？

链和反链



宽度为9，比刚才的大，那么这个是最大吗?

链和反链

- 观察:
  - 如果一个poset P可以划分为t个anti-chains，即
    $\exists$anti-chains $A_1, A_2, ... A_t$ with $P = \cup_{i=1}^{t} A_i$， 那么显然P的高度最多为t
    - 因为一旦某个链C长超过t, 根据鸽巢原理，那么必然C和A_i至少有两个交集，矛盾!
  - 如果一个poset P可以划分为s个chains，即
    $\exists$chains $C_1, C_2, ... C_s$ with $P = \cup_{i=1}^{s} C_i$， 那么显然P的宽度最多为s
    - 同样，根据鸽巢原理，可以知道最大的反链也不会超过s个元素

链和反链

相等！

Robert P. Dilworth

Theorem (Dilworth, 1950). A poset of width w can be partitioned into w chains

Dilworth，Fulkerson等人早就发现了这个结论，但觉得有些平凡，没有发表

Leonid Mirsky

Theorem (Mirsky's Theorem, Dual Dilworth, 1971). A poset of height h can be partitioned into h anti-chains

D. R. Fulkerson

Fulkerson (1954) Used bipartite matching algorithm (network flows) to find minimum chain partition and maximum anti-chain simultaneously.

偏序集的一些操作

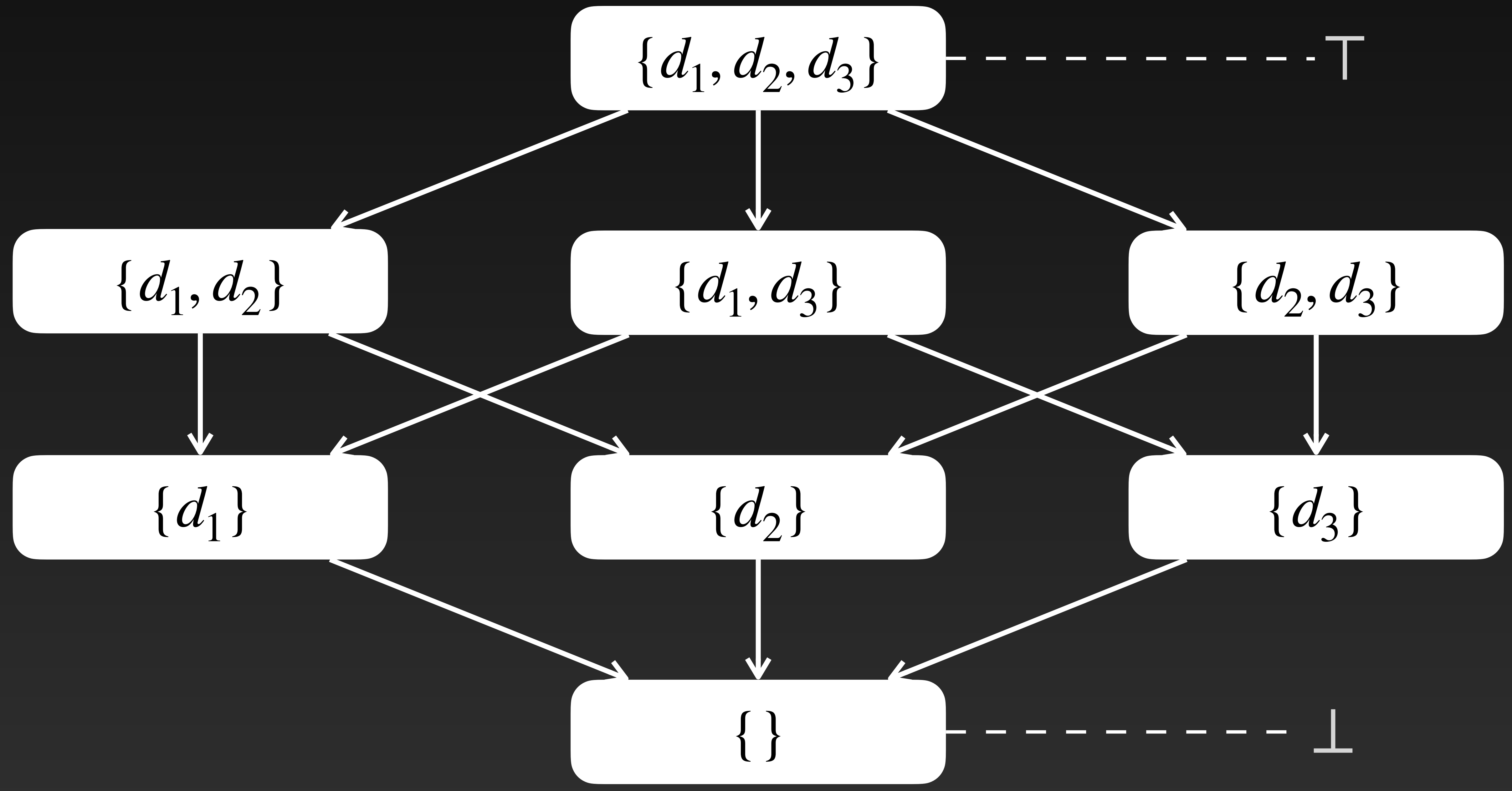
- 给定两个posets P, Q:
  - cardinal sum:
    - $P + Q := (P \cup Q, \preceq_{P+Q})$, s.t., $x \preceq_{P+Q} y \iff x \preceq_P y$ or $x \preceq_Q y$
  - cardinal product:
    - $P \times Q := (P \times Q, \preceq_{P\times Q})$, s.t., $(x, y) \preceq_{P\times Q} (x', y') \iff x \preceq_P x'$ and $y \preceq_Q y'$
    - P乘以自己n次记为P^n

格（Lattice）

- 可以刻画很多计算（比如布尔代数）
  - 定义：一个格$(L, \preceq, \wedge, \vee)$是一个偏序集$(L, \preceq)$, 其中每一对元素都有一个最大下界（greatest lower bound, glb）和一个最小上界(least upper bound, lub)
  - 由于lub和glb的唯一性，二元操作符∨(join)和∧(meet)可以这样定义：
    - $a \vee b = \text{lub}(a, b)$
    - $a \wedge b = \text{glb}(a, b)$
  - 每一个拥有有限元素的格L（non-empty）都有一个最小元素（最小元，⊥）和一个最大元素（最大元，⊤), 使得对于一个元素$x \in L$而言，$x \preceq \top$， $\bot \preceq x$ （可以通过数学归纳证明）

格的例子

- 令$V = \left\{x \mid x \subseteq \{d_1, d_2, d_3\}\right\}$, $\wedge = \cap$, $\vee = \cup$)



- 最高和最低元素
  - 最大元$\top$, 使得对任何x, 都有$x \wedge \top = x$, 容易看出： $\top = \{d_1, d_2, d_3\}$
  - 最小元$\perp$, 使得对于任何x, 都有$x \wedge \perp = \perp$， 容易看出： $\perp = \{\}$

半格（Semi-Lattice）

- 给定一个poset $(P, \preceq)$
  - 我们称$(P, \preceq, \vee)$为一个并半格（join semi-lattice，upper semi-lattice）,当$\forall x, y \in P$, x, y的最小上界存在，即$x \vee y$存在
  - 我们称$(P, \preceq, \wedge)$为一个交半格（meet semi-lattice，lower semi-lattice）,当$\forall x, y \in P$, x, y的最大下界存在，即$x \wedge y$存在

半格

Theorem 1 A join semi-lattice $(P, \preceq, \vee)$ satisfies:

$(a \vee b) \vee c = a \vee (b \vee c)$	结合律(associativity)
$a \vee b = b \vee a$	交换律(commutativity)
$a = a \vee a$	幂等性(idempotence)

- Theorem 1.
- 结合律： let $m = (a \vee b) \vee c$, we have $a \vee b \preceq m, c \preceq m$. Therefore, $a \preceq m, b \preceq m, c \preceq m$. Further, $a \preceq m, (b \vee c) \preceq m$. As a result, $a \vee (b \vee c) \preceq m$. Similarly, we can have $(a \vee b) \vee c \preceq n$, where $n = a \vee (b \vee c)$. At last, $m = n$
- 交换律： 证法类似上述过程
- 幂等性：注意到$a \preceq a$, 其余类似

Theorem 2 A meet semi-lattice $(P, \preceq, \wedge)$ satisfies:

$(a \wedge b) \wedge c = a \wedge (b \wedge c)$	结合律(associativity)
$a \wedge b = b \wedge a$	交换律(commutativity)
$a = a \wedge a$	幂等性(idempotence)

- Theorem 2.
- let $m = (a \wedge b) \wedge c$, we have $m \preceq (a \wedge b), m \preceq c$. Therefore, $m \preceq a, m \preceq b, m \preceq c$.
- 其余证法类似

半格的等价定义

并半格的另一个角度

Theorem Let X be a set with function $\vee: X \times X \to X$ satisfying:

$(a \vee b) \vee c = a \vee (b \vee c)$ 结合律(associativity)

$a \vee b = b \vee a$ 交换律(commutativity)

$a = a \vee a$ 幂等性(idempotence)

Let $a \preceq b \triangleq (a \vee b) = b$. Then, $(X, \preceq, \vee)$ is a join semi-lattice

- 首先，$(X, \preceq)$是偏续集：
  - 传递性（Transitive）:如果 $x \preceq y$并且$y \preceq z$, 那么$x \preceq z$
    - $x \vee y = y, y \vee z = z$, 那么$x \vee z = x \vee (y \vee z) = (x \vee y) \vee z = y \vee z = z$ (结合)
  - 反对称性 (Anti-symmetric): 如果如果 $x \preceq y$并且$y \preceq x$, 那么$x = y$
    - $x \wedge y = y, y \wedge x = x$, 那么 $x = y \wedge x = x \wedge y = y$ (交换)
  - 自反性（Reflexive）: $x \preceq x$
    - $x \vee x = x$ (幂等)

半格的等价定义

并半格的另一个角度

Theorem Let X be a set with function $\vee : X \times X \to X$ satisfying:

$(a \vee b) \vee c = a \vee (b \vee c)$ 结合律(associativity)

$a \vee b = b \vee a$ 交换律(commutativity)

$a = a \vee a$ 幂等性(idempotence)

Let $a \preceq b \triangleq (a \vee b) = b$. Then, $(X, \preceq, \vee)$ is a join semi-lattice

- 其次，∨是lub：
  - 因为 $a \vee (a \vee b) = (a \vee a) \vee b = a \vee b$， 我们有$a \preceq a \vee b$, 同理我们有$b \preceq a \vee b$
  - 令$a \preceq x$ and $b \preceq x$. 我们有

    $$a \vee x = x = b \vee x \implies (a \vee (b \vee x)) = x \implies ((a \vee b) \vee x) = x \implies (a \vee b) \preceq x$$
  - 有上述两点可知：∨是lub

半格的等价定义

交半格的另一个角度

Theorem Let X be a set with function $\wedge: X \times X \to X$ satisfying:

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \quad \text{结合律(associativity)}$$

$$a \wedge b = b \wedge a \quad \text{交换律(commutativity)}$$

$$a = a \wedge a \quad \text{幂等性(idempotence)}$$

Let $a \preceq b \triangleq (a \wedge b) = a$. Then, $(X, \preceq, \wedge)$ is a meet semi-lattice

- 证明类似

格的性质

Theorem A lattice $(P, \preceq, \vee, \wedge)$ satisfies:

$$
\begin{aligned}
(a \vee b) \vee c &= a \vee (b \vee c) && \text{结合律(associativity)} \\
(a \wedge b) \wedge c &= a \wedge (b \wedge c) && \\
a \vee b &= b \vee a && \text{交换律(commutativity)} \\
a \wedge b &= b \wedge a && \\
a &= a \vee a && \text{幂等性(idempotence)} \\
a &= a \wedge a && \\
a \wedge (a \vee b) &= a && \text{吸收律(absorption)} \\
b \vee (a \wedge b) &= b &&
\end{aligned}
$$

- 吸收律的证明：
- $a \preceq a, a \preceq (a \vee b) \Longrightarrow a$ 是一个a和$a \vee b$的下界.
- 令$x \preceq a$ and $x \preceq a \vee b$. 我们有x是a和$a \vee b$的下界. 显然$x \preceq a$. 因此， a是最小下界，即$a \wedge (a \vee b) = a$
- $a \wedge b \preceq b, b \preceq b \Longrightarrow b$ 是一个$a \wedge b$和b的上界， 同样任何$a \wedge b$和b的上界都必然比b大，因此$b \vee (a \wedge b) = b$

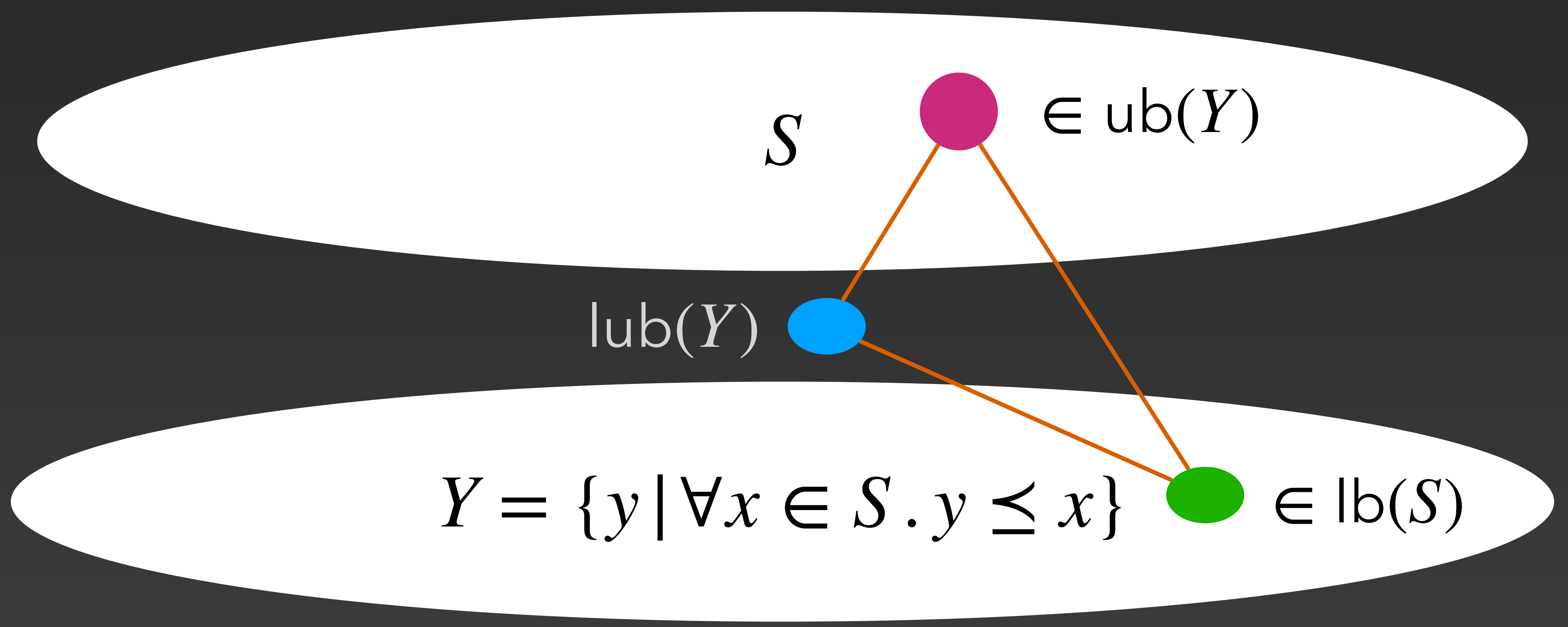
完备格(Complete Lattice)

条件比任意两个元素（格）要强，因为集合可能有无穷的元素

- 完全格$(X, \preceq)$是一个偏序集，对于其每个子集，都有一个最小上界和最大下界，即 $\forall S \subseteq X, \exists \text{lub}(S) \in X$ and $\exists \text{glb}(S) \in X$
  - 实际上，对于一个偏序集$(X, \preceq)$而言，lub(S)和glb(S)只需存在一个即是完备格

Theorem let $(X, \preceq)$ be a poset satisfying $\forall S \subseteq X, \text{lub}(S) \in X$. Then $(X, \preceq)$ is a complete lattice. Similarly, if a poset $(X, \preceq)$ satisfies $\forall S \subseteq X, \text{glb}(S) \in X$. Then $(X, \preceq)$ is a complete lattice.

如果$(X, \preceq)$满足$\forall S \subseteq X, \text{lub}(S) \in X$. 可以得出对于一个给定的子集$S$, 那么其最大下界glb是$S$所有下界的最小上界，即 $\text{glb}(S) = \text{lub}\{y \mid \forall x \in S . y \preceq x\}$



完备格(Complete Lattice)

- 完备格具备的一些性质：

Theorem a complete lattice $(X, \preceq)$ satisfies

1. It has a least element $\perp$
2. It has a greatest element $\top$

显然地，因为$X \subseteq X$, 因此 glb$(X) \in X$, lub$(X) \in X$， 即其拥有最小元和最大元

完备格(Complete Lattice)

- 几个显然的完备格：
  - 任何非空的有限格都是完备格
    - 数学归纳法（在格的大小上）
  - 任何格$(X, \preceq, \vee, \wedge)$拥有Ascending Chain Condition (no infinite strictly ascending sequence, $a_1 \prec a_2 \prec a_3 \ldots$, where $a_i \in X$), 并且有一个最小元$\perp$, 那么其就是一个完备格
    - 哪怕是拥有无限元素的格，只要有一个最小元，并且没有无限上升链，那么其任何子集都会有一个最小上确界

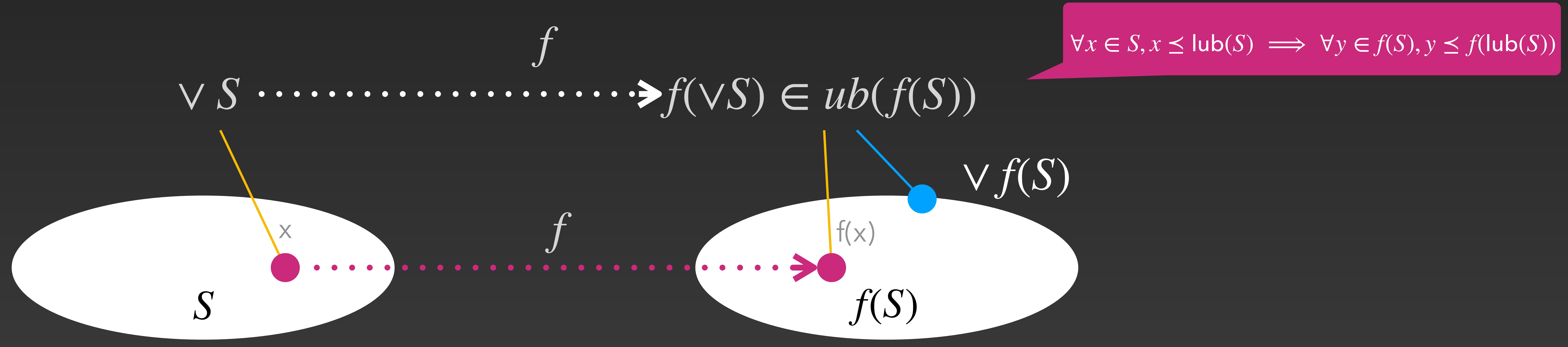
乘积格（Product Lattice）

- 对多个格进行笛卡尔积操作，形成新的格，即：
  - 给定格$(P_1, \preceq_1), (P_2, \preceq_2), \ldots, (P_n, \preceq_n)$， 那么$(P_1, \preceq_1) \times (P_2, \preceq_2) \times \ldots \times (P_n, \preceq_n)$也是一个格
    - 其中数据形式为 $< e_1, e_2, \ldots, e_n >$, where $e_i \in P_i$
    - $< e_1, e_2, \ldots, e_n > \preceq < e'_1, e'_2, \ldots, e'_n >$ 当且仅当$\forall e_i, e'_i, e_i \preceq_i e'_i$
    - $< e_1, e_2, \ldots, e_n > \wedge < e'_1, e'_2, \ldots, e'_n > = < e_1 \wedge e'_1, e_2 \wedge e'_2, \ldots, e_n \wedge e'_n >$
    - $< e_1, e_2, \ldots, e_n > \vee < e'_1, e'_2, \ldots, e'_n > = < e_1 \vee e'_1, e_2 \vee e'_2, \ldots, e_n \vee e'_n >$
  - 如果对所有的$(P_i, \preceq_i)$都是一个完全格，那么$(P_1, \preceq_1) \times (P_2, \preceq_2) \times \ldots \times (P_n, \preceq_n)$也是一个完全格
    - 结论显然. 可以看到新的top元素为： $< \top_1 , \top_2 , \ldots , \top_n >$,新的bot元素为 $< \perp 1, \perp 2, \ldots, \perp n >$

单调映射

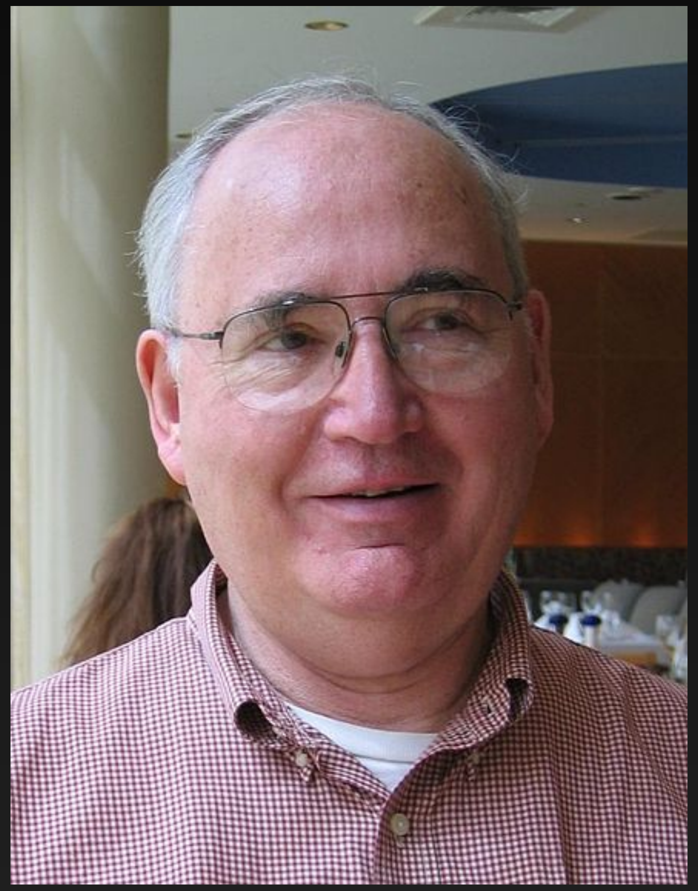
- 对于个偏续集$(X, \preceq_X), (Y, \preceq_Y)$，而言$f: X \rightarrow Y$是一个单调映射（monotone map, alternatively, order-preserving）当$\forall x, y \in X . x \preceq_X y \implies f(x) \preceq_Y f(y)$

Theorem For posets $(X, \preceq_X), (Y, \preceq_Y)$, let $f: X \rightarrow Y$ be monotone map. For $S \subseteq X$, assume $\vee S$, and $\vee f(S)$ exist, then $\vee f(S) \preceq_Y f(\vee S)$



连续映射

- 对于偏序集$(X, \preceq_X), (Y, \preceq_Y)$而言，$f: X \to Y$是连续的（continuous, actually, upper-continuous, 也被称为Scott-continuous），当且仅当其保持least upper bounds性质 (也称limit preserving)，即对于X中的任何一个链$C \subseteq X$, 如果$\vee C$存在，那么$\vee f(C)$存在，且$f(\vee C) = \vee f(C)$. 类似地，我们可以定义lower-continuous


Dana Stewart Scott

Theorem If $x \preceq y$, and f is continuous then $f(y) = f(x) \vee f(y)$

y is $\vee \{x, y\}$, $f(x) \vee f(y)$ is $\vee (\{f(x), f(y)\})$

Theorem Continuous maps are monotonic

$f(y) = f(x) \vee f(y) \implies f(x) \preceq f(y)$

Theorem Let poset $(X, \preceq_X)$ satisfies ascending chain condition(ACC), let $(Y, \preceq_Y)$ be a poset. Let $f: X \to Y$ be a monotone function. Then, f is continuous.

映射复合

Theorem Any composition of monotone function is monotone:
Given posets $(X, \preceq_X)$, $(Y, \preceq_Y)$,$(Z, \preceq_Z)$, let $f : X \to Y, g : Y \to Z$ both be monotone, then $h = g \circ f : X \to Z$ is monotone.

Theorem Any composition of continuous function is monotone:
Given posets $(X, \preceq_X)$, $(Y, \preceq_Y)$,$(Z, \preceq_Z)$, let $f : X \to Y, g : Y \to Z$ both be continuous, then $h = g \circ f : X \to Z$ is continuous.

不动点(fixed point)

- 给定X为一个集合，一个操作$f: X \to X$的不动点即为X的一个元素x，使的$f(x) = x$
- 令f是偏续集$(X, \preceq)$上的一个操作
  - $fp(f) \triangleq \{x \mid f(x) = x\}$
  - $prefp(f) \triangleq \{x \mid f(x) \preceq x\}$
  - $postfp(f) \triangleq \{x \mid x \preceq f(x)\}$

$\longrightarrow fp(f) = prefp(f) \cap postfp(f)$

  - least fixed point $lfp(p) \triangleq \min(fp(f))$
  - greatest fixed point $gfp \triangleq \max(fp(f))$

克纳斯特-塔斯基定理(Knaster-Tarski fixed point theorem)

Theorem A monotonic map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$ has a least fixed point and a greatest fixed point, which are：

1. $lfp(f) = \wedge\, prefp(f) = \wedge\, \{x \mid f(x) \preceq x\}$
2. $gfp(f) = \vee\, postfp(f) = \vee\, \{x \mid x \preceq f(x)\}$
3. Moreover, the fixed points form a complete lattice

Bronisław Knaster

Alfred Tarski

- 证明：1. 令 $prefp(f)$ 为f所有的prefixed points， 令p为$prefp(f)$的glb，即$p = \wedge\, prefp(f)$. 首先由于X是完备格，p必然存在。下面我们来证明p也同时是一个 a）least prefixed point和 b）least fixed point
  - a）对于任何的prefixed point x而言:
    - 我们有$p \preceq x$， 因此$f(p) \preceq f(x)$（f是单调的）
    - 由于x是一个prefixed point，因此$f(x) \preceq x$， 因此$f(p) \preceq f(x) \preceq x \implies f(p) \preceq x$.
    - 至此，$f(p)$是$prefp(f)$的一个下界. 由于p是$prefp(f)$的最大下界，我们有 $f(p) \preceq p$, 意味着p本身是一个prefixed point.
    - 因此，p是一个prefixed point, 也是$prefp(f)$的一个下界， 我们有 p是least prefixed point, 即 $\wedge\,(prefp(f))$就是prefixed point集合的最小元

克纳斯特-塔斯基定理(Knaster-Tarski fixed point theorem)

Theorem A monotonic map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$ has a least fixed point and a greatest fixed point, which are：

1. $lfp(f) = \wedge\, prefp(f) = \wedge\, \{x \mid f(x) \preceq x\}$
2. $gfp(f) = \vee\, postfp(f) = \vee\, \{x \mid x \preceq f(x)\}$
3. Moreover, the fixed points form a complete lattice

Bronisław Knaster

Alfred Tarski

- 证明： b）由于p是一个prefixed point.
    - 我们有$f(p) \preceq p$. 又f是单调的，我们有$f(f(p)) \preceq f(p)$
        - 因此$f(p)$也是一个prefixed point. 由于p本身是prefixed points的下界，我们有$p \preceq f(p)$
    - 至此，我们有$p \preceq f(p), f(p) \preceq p$, 因此poset的反传递性，我们有$p = f(p)$
    - 因此，p是一个不动点
    - 再注意到，所有的不动点本身都是prefixed points. 而p又是所有prefixed points的下界，所以p是所有不动点的下界
    - 至此，p就是最小的不动点

克纳斯特-塔斯基定理(Knaster-Tarski fixed point theorem)

Theorem A monotonic map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$ has a least fixed point and a greatest fixed point, which are:

1. $lfp(f) = \wedge\, prefp(f) = \wedge \{x \mid f(x) \preceq x\}$
2. $gfp(f) = \vee\, postfp(f) = \vee \{x \mid x \preceq f(x)\}$
3. Moreover, the fixed points form a complete lattice


Bronisław Knaster


Alfred Tarski

- 证明：2的证明和1类似（对偶），至此省略

克纳斯特-塔斯基定理(Knaster-Tarski fixed point theorem)

Theorem A monotonic map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$ has a least fixed point and a greatest fixed point, which are：

1. $lfp(f) = \wedge\, prefp(f) = \wedge\, \{x \mid f(x) \preceq x\}$
2. $gfp(f) = \vee\, postfp(f) = \vee\, \{x \mid x \preceq f(x)\}$
3. Moreover, the fixed points form a complete lattice

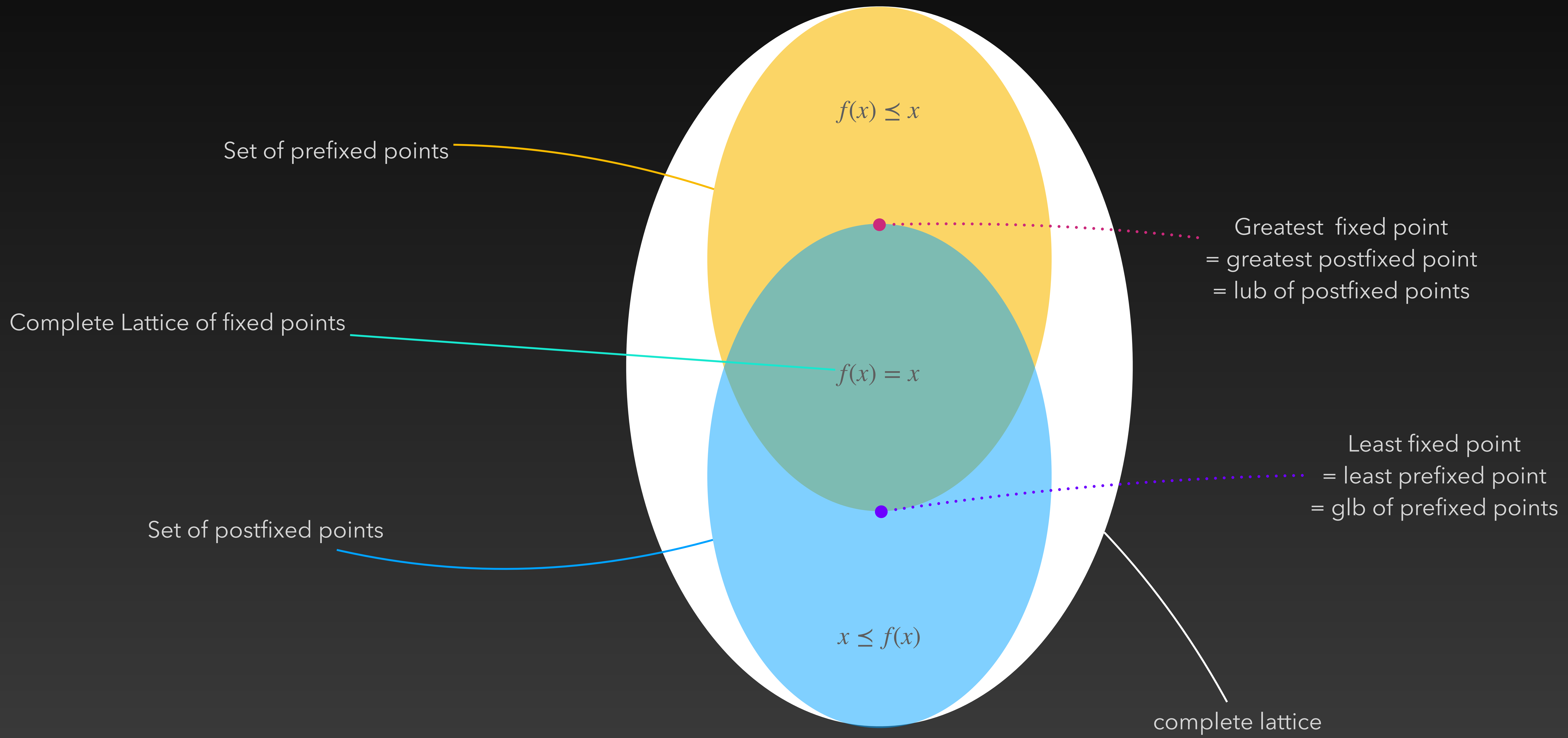

Bronisław Knaster


Alfred Tarski

- 证明：3. 令W为fixed points的一个子集，我们来证明其存在一个上确界(lub)在fixed points集中. 对偶地，如果我们能证明其还存在一个下确界(glb)就可以证明fixed points构成了一个完备格.
  - 令$q = \text{lub}(W)$ (注意：这是从X的角度看，$q \in X$)，$\hat{W} = \{w \mid q \preceq w\}$. 我们有$q \in \hat{W}, q = \text{glb}(\hat{W})$
  - $\hat{W}$作为一个完备格的子集，有一个lub和glb. 我们有$q = \text{glb}(\hat{W}), q \in \hat{W}$, 因此$\text{glb}(\hat{W}) \in \hat{W}$. 进一步，由于$q \in \hat{W}, q \preceq \text{lub}(\hat{W})$. 因此根据$\hat{W}$的定义，$\text{lub}(\hat{W}) \in \hat{W}$. 由于$\text{glb}(\hat{W}) \in \hat{W}$，$\text{lub}(\hat{W}) \in \hat{W}$， $\hat{W}$显然是一个完备格（any subset has lub and glb in it)
  - f从$\hat{W}$映射到$\hat{W}$自身： 对于W的任意元素w, 和$\hat{W}$的任意元素x，我们有$w \preceq q, q \preceq x$, 因此由f的单调性，$f(w) \preceq f(q) \preceq f(x)$
    - W为fixed points的一个集合，因此$f(w) = w$, 意味着$w \preceq f(x)$. 由w的任意性，$\text{lub}(W) \preceq f(x)$. 再因为$q = \text{lub}(W)$, 我们有$q \preceq f(x)$
    - 根据$\hat{W}$的定义，$f(x) \in \hat{W}$
  - 基于上述两点，f是一个完备格$\hat{W}$上的一个映射(其本身是单调的). 因此，其在$\hat{W}$上含有一个最小不动点$\hat{q}$ （a.k.a. $f(\hat{q}) = \hat{q}$）. 由于q是$\hat{W}$上的最小元，我们有$q \preceq \hat{q}$. 因此，如果我们从fixed points集合来看，$\hat{q}$就是W的一个上确界

克纳斯特-塔斯基定理(Knaster-Tarski fixed point theorem)

定理的一个图示

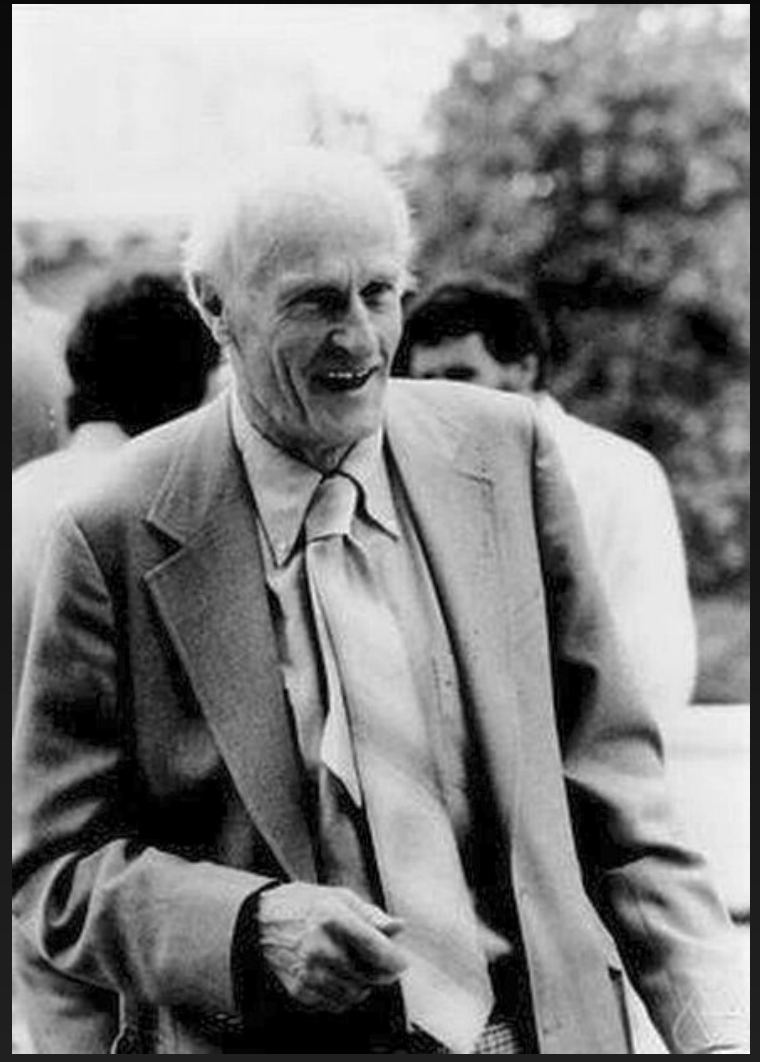


克莱尼不动点定理(Kleene's fixed point theorem)

- Knaster-Tarski's theorem说明了lfp和gfp的存在（当然，也存在相应的算法，但是不那么直接）. 克莱尼不动点定理则给出了一种计算（或者近似计算，在continous性质不满足的情况下）lfp和gfp.

Theorem Let a monotonic map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$. Then $f^{\alpha}(\bot) \preceq f^{\alpha+1}(\bot)$ for all ordinals α. That is, $\bot \preceq f(\bot) \preceq f^2(\bot) \ldots f^i(\bot)$ is a chain

By induction.
And the base case
$\bot \preceq f(\bot)$ always holds



Stephen Kleene

Theorem (Kleene) Let a continuous map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$. Then $\vee \{f^i(\bot) \mid n \in \mathbb{N}\}$ is the least fixed point of f.

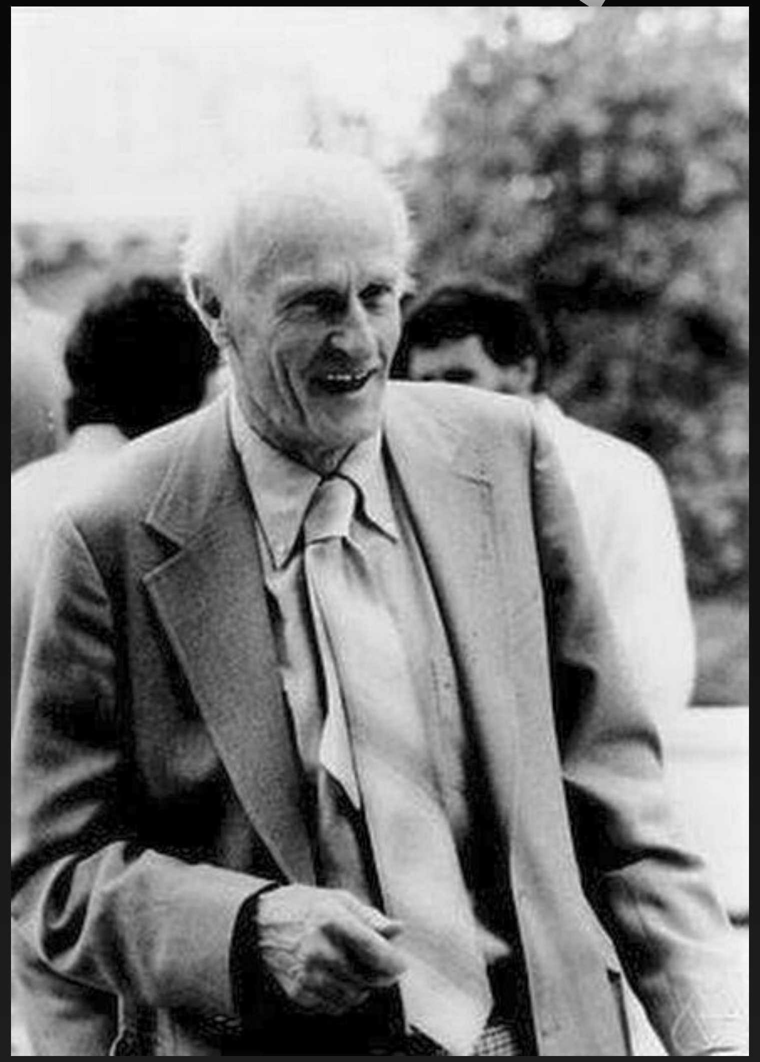
$$lfp(f) = \vee_{n\in\mathbb{N}} f^i(\bot)$$

- We first prove that $lfp(f)$ is a fixed point:
  - $f(lfp(f)) = f(\vee_{n\in\mathbb{N}} f^i(\bot)) \overset{\text{by continuity of } f}{=} \vee_{n\in\mathbb{N}} f(f^i(\bot)) \overset{\text{by def of } f^n}{=} \vee_{n\in\mathbb{N}} f^{i+1}(\bot) \overset{\text{Note: we only add } \bot\text{, but } \forall x, x \vee \bot = x}{=} \vee_{n\in\mathbb{N}} f^i(\bot) = lfp(f)$

克莱尼不动点定理(Kleene's fixed point theorem)

Theorem (Kleene) Let a continuous map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$. Then $\vee \{f^i(\bot) \mid n \in \mathbb{N}\}$ is the least fixed point of f.

$$lfp(f) = \vee_{n\in\mathbb{N}} f^i(\bot)$$



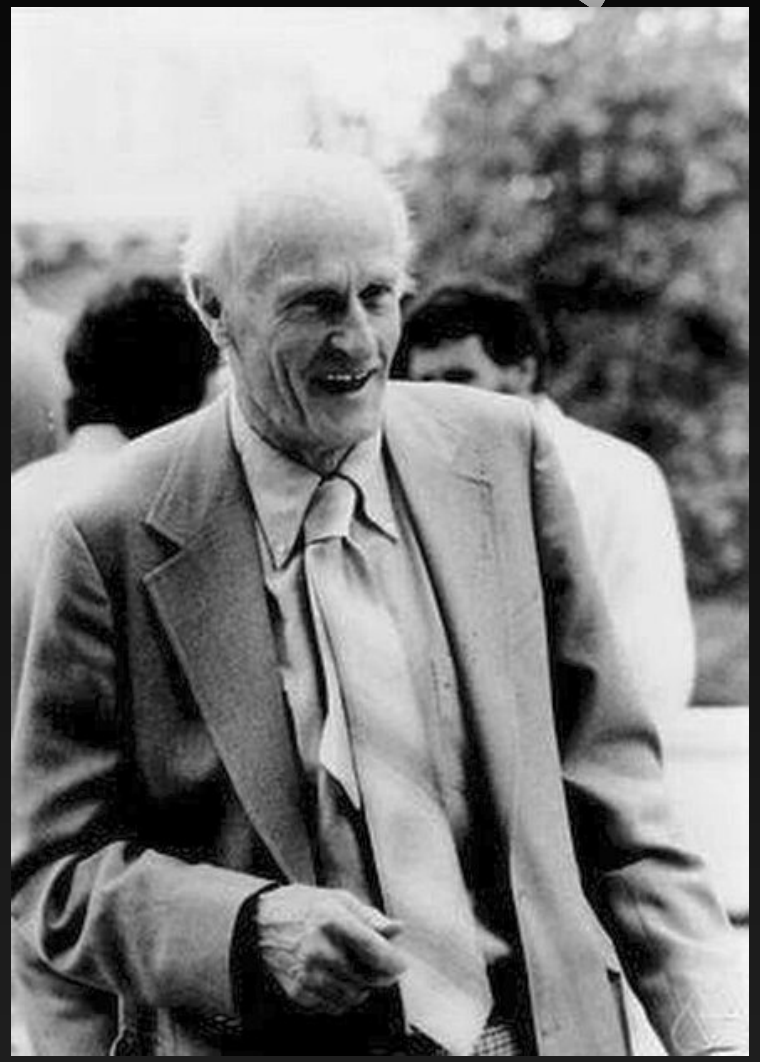
Stephen Kleene

- We then prove that $lfp(f)$ is the least:
  - We only need to prove $\forall x \in X . f(x) \preceq x \implies lfp(f) \preceq x$
    - Since any fixed point is a prefixed point.
  - We then only need to prove $\forall i \in \mathbb{N} . f^i(\bot) \preceq x$, then we get x is ub of this chain (also refer to Kleene chain)
  - Induction: $f^0(\bot) = \bot \preceq x$
  - Assume $f^n(\bot) \preceq x$, then
    - $f^{n+1}(\bot) = f(f^n(\bot)) \underbrace{\preceq}_{\text{Monotone of } f} f(x) \preceq x$

克莱尼不动点定理(Kleene's fixed point theorem)

Theorem (Kleene) Let a continuous map $f : X \to X$ on a complete lattice $(X, \preceq, \top, \bot, \wedge, \vee)$. Then $\vee \{f^i(\bot) \mid n \in \mathbb{N}\}$ is the least fixed point of f.

$$lfp(f) = \vee_{n\in\mathbb{N}} f^i(\bot)$$



Stephen Kleene

- 根据kleene不动点定理，我们只要从⊥出发，不断迭代f, 就可以找到f的不动点，而且是最小的不动点
- 此外，我们可以对偶地得到最大不动点：
  - $gfp(f) = \wedge_{i\in\mathbb{N}} f^i(\top)$
  - 即从⊤出发，不断迭代f, 就可以找到f的最大不动点

回到数据流分析

回顾之前的两种分析

	Reaching Definitions	Live Varaibles
Domain	Sets of definitions	Sets of variables
Direction	Forward: $OUT[b] = f_b(IN[b])$ $IN[b] = \wedge\ OUT[pred(b)]$	Backward: $IN[b] = f_b(OUT[b])$ $OUT[b] = \wedge\ IN[succ(b)]$
Transfer function	$f_b(IN[b]) = Gen(b) \cup (IN[b] - Kill(b))$	$f_b(OUT[b]) = Use(b) \cup (OUT[b] - Def(b))$
Meet operation （∧）	U	U
Boundary condition	$OUT[entry] = \varnothing$	$IN[exit] = \varnothing$
Initial interior points	$OUT[b] = \varnothing$	$IN[b] = \varnothing$

单调数据流框架（Monotone Dataflow Frameworks）

- 一个统一的数据流分析框架，由Gary A. Kildall (POPL 1973)提出 Jon B. Kam 和 Jeffery D. Ullman (Acta Info 1977)拓展，因此也被称为Kildall数据流框架
- 该框架是一个四元式 $(G, \mathscr{L}, F, D)$
  - G 是一个控制流图 $< N, E, n_0 >$ ， $N = n_0, n_1, n_2, \ldots$ 是基本块，E是基本块之间的控制流边，n_0是开始节点
  - $\mathscr{L}$是一个高度有限的格(即为完备格，其实semi-lattice也一样，有限意味着其必然符合ACC要求)，一个meet操作∧和join操作∨, 有一个最高T和最低元素⊥，即$(L, \preceq, \vee, \wedge, \top, \bot)$
    - $\mathscr{L} = L_0 \times L_1 \times L_2 \times \ldots$ ，其中L_i 代表了基本块节点n_i的可能的解（值的域），L是这些格的积，代表了整个程序的解。一般而言,每个格的解空间是相同的，比如都是程序中的n个变量的可达定义问题，因此整体的格就可以表达为L^n, 每个节点都是L
  - F 是一个函数空间, $\forall f \in F, f : L \to L$

    如果每个函数都monotone （前面的ACC 必然导致也是continuous）, 那么就有不动点解了！
    - 存在一个全等函数f，对于所有的x，都有$f(x) = x$ （用于对应空block）
    - 对于复合闭包，如果$f_1, f_2 \in F$, 那么$f_1 \circ f_2 \in F$ （这样多个block的transfer function就可以复合）
  - D 分析方向, 前向还是后向

例子

- 定义可达分析：
  - 对于程序中的每个基本块，L为definitions的可达集合，即2^{Defs}， ⊤ = all the varaibles，⊥ = { }，∧ = ∩ ，∨ = ∪， ≤ = ⊆
    - 整体的格是笛卡尔积$L^n = 2^{Defs} \times 2^{Defs} \times 2^{Defs}$
  - 对于基本块而言$f : L \to L$, $f(x)$ = Gen ∪ (x − Kill)
    - $x_1 \le x_2 \implies$ Gen ∪ (x_1 − Kill) ≤ Gen ∪ (x_2 − Kill)
    - Alternatively （等价的表述）， Gen $\bigcup$ ($x_1 \cup x_2$ − Kill) ≤ Gen ∪ (x_1 − Kill) $\bigcup$ Gen ∪ (x_2 − Kill)
  - 另一方面，对于控制流边也可以看出是一种function: $L \times L \to L$, $(x_1, x_2) \le (x_1', x_2') \implies x_1 \cup x_2 \le x_1' \cup x_2'$
  - 因此，这个框架中的每个function都是单调的 (由于格为有限格，因此也是连续的)
  - 方向 D = forward

例子

- 定义可达分析：
  - 整体的域是一个完备格L^n
  - 所有的基本块函数、控制流函数本质上是对程序值域的一个约束方程：
    - 每一个$f: L \to L$，都是表达了输入一个x，输出一个$f(x)$

The merge can be regarded as $f: L \times L \to L$

我们可以把值域用笛卡尔积统一，因此所有的函数都是$f_i: L^n \to L^n$, 而程序的所表达的整体的语义可以看成是这些函数的复合$f \triangleq f_1 \circ f_2 \ldots$，其必然也是monotone的

例子

Abstract domain

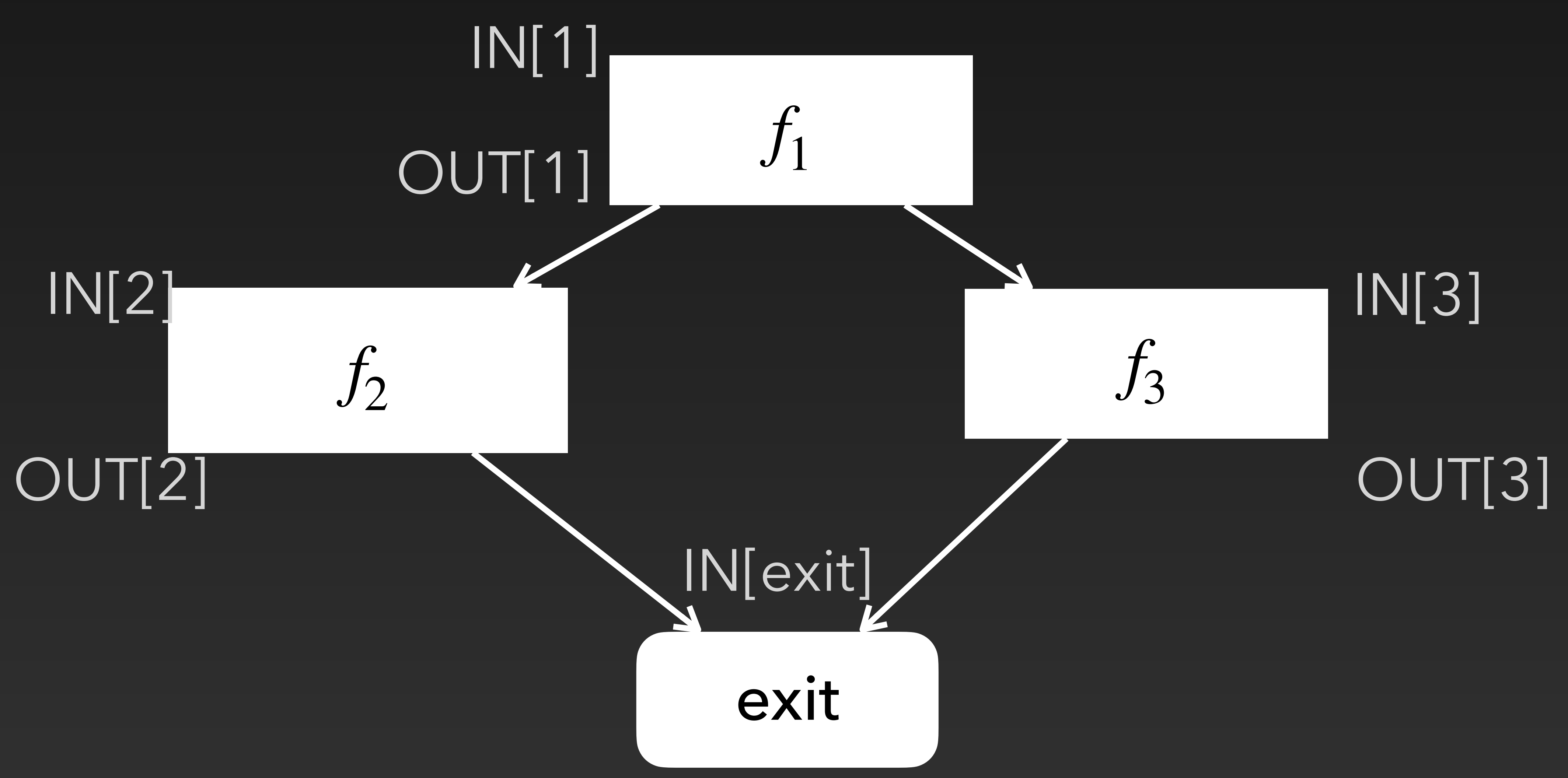
Meet or join

Fixed point

- 定义可达分析：
  - 我们想要的是：无论程序的输入是什么，沿着什么样的path，经过多少循环，我们得到的关于这个的数据信息的判定都为真
    - 因此我们需要关于复合的约束方程f，得到其不动点：即一定为真的信息！
      - 可以类比为循环不变式
    - 进一步，我们其实想要的是最小/最大的不动点
      - 可以类比为有很多循环不变式（都为真），但有些是平凡的（没多大用），那些最强/最弱的不变式才能证明我们所需要的性质
      - 比如对于冒泡排序而言，每次迭代元素不变必然为真（但没什么用），但是每次迭代前i个元素有序则非常有用，并且 每次迭代前i个元素有序 $\Longrightarrow$ 每次迭代元素不变 (但反之不然)，这其实和我们之前的所有不动点形成了一个完备格类似！

例子

- 对定义可达而言，定义可达的越多越没用（利用哑节点来detect undefined varaibles use），但越安全（越可能为真），定义越少越有用，但越不安全（可能不为真）



(⊤ , ⊤ , ⊤ , ... ⊤) 必然为真，但没什么用

(⊥ , ⊥ , ⊥ , ... ⊥) 最为激进（精确），但最可能不安全

因此，对于这个问题，我们希望找到最小不动点 lfp！为真，而且最为精确（least！）

因此，本质上我们是对一个完备格L^n，上关于约束的方程f（所有基本块的transfer function，path的merge function的复合）的最小不动点. 由于L^n高度有限，f是单调的也是连续的，因此根据Kleene不动点定理必然能找到最小不动点！

算法的复杂度

- Kleene不动点定理的一个具体实现就是依次迭代（apply 每个节点和汇聚的f）
  - 那么大概要迭代多少次才能找到最小不动点呢?
    - 其实就是完备格L^n的高度 （因为每次迭代都是在完备格上做的一次≤的迁移，即Kleene chain上的移动）
  - 能够优化这个算法吗?
    - worklist算法时可以根据数据依赖图来进行迭代（因为有些点的数据改变并不会改变另一个节点数据信息）

另一个例子：常量传播(Constant Propagation)

- 常量传播： 将表达式中能够确定每次求值都为某个常量（evaluated to be some constant）的表达式替换为这个常量
  - 常见的优化技术，可以减少运行时求值需要的时间
- 值域：Undef, ..., -2, -1, 0, 1, 2, ..., NAC
  - NAC: Not A Constant， 对于一个变量有多个值定义点，因此不是某个常量
  - Undef ： undefined，对于一个变量的值没有定义

另一个例子：常量传播(Constant Propagation)

- 聚合∧的作用是汇聚多个路径中的定义，在常量传播中的定义如下

一个变量在多个路径上都是相同的值，就是这个常量，否则就是一个可能随着执行不同的变量

x	y	x ∨ y	
c1	c2	NAC	c1 != c2
c1	c2	c1	c1 = c2
undef	c1	c1	
undef	undef	undef	
undef	NAC	NAC	
NAC	c1	NAC	
NAC	undef	NAC	
NAC	NAC	NAC	

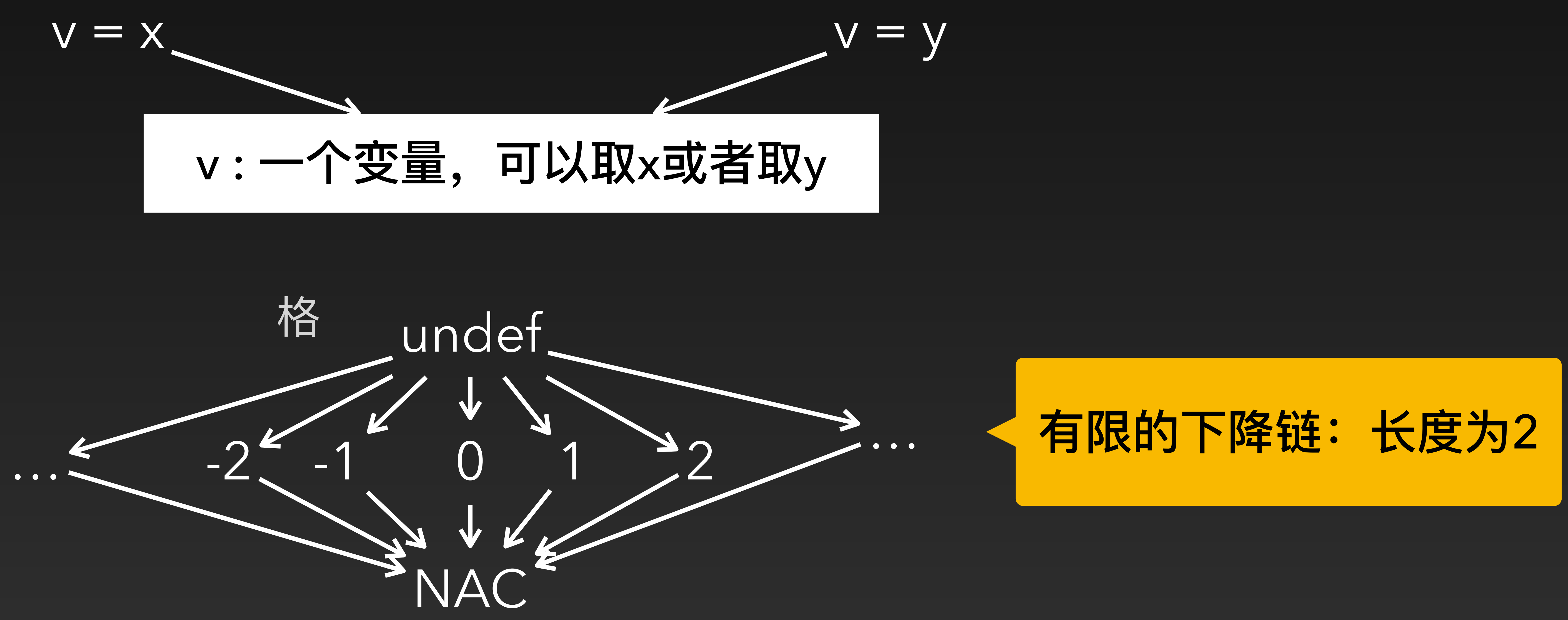
如果某条路径是未定义，说明在这个路径上，这个变量的值is not interesting， 即其可以为任何值也不会影响程序。那么如果另外一个路径上其定义为常量，对于优化而言，将其设为这个常量是安全的。

注意：变量未定义就使用（use before define）本质上是一个未定义行为，可能会有隐患。但作为常量传播的优化而言，无需为这个未定义行为负责。此外，也可以看出，一些优化技术甚至可以消除未定义行为！

只要某个路径上不是一个常量（存在多个可能的值定义），那么汇聚后一定不是一个常量

另一个例子：常量传播

- 聚合v给出了一个如下的格

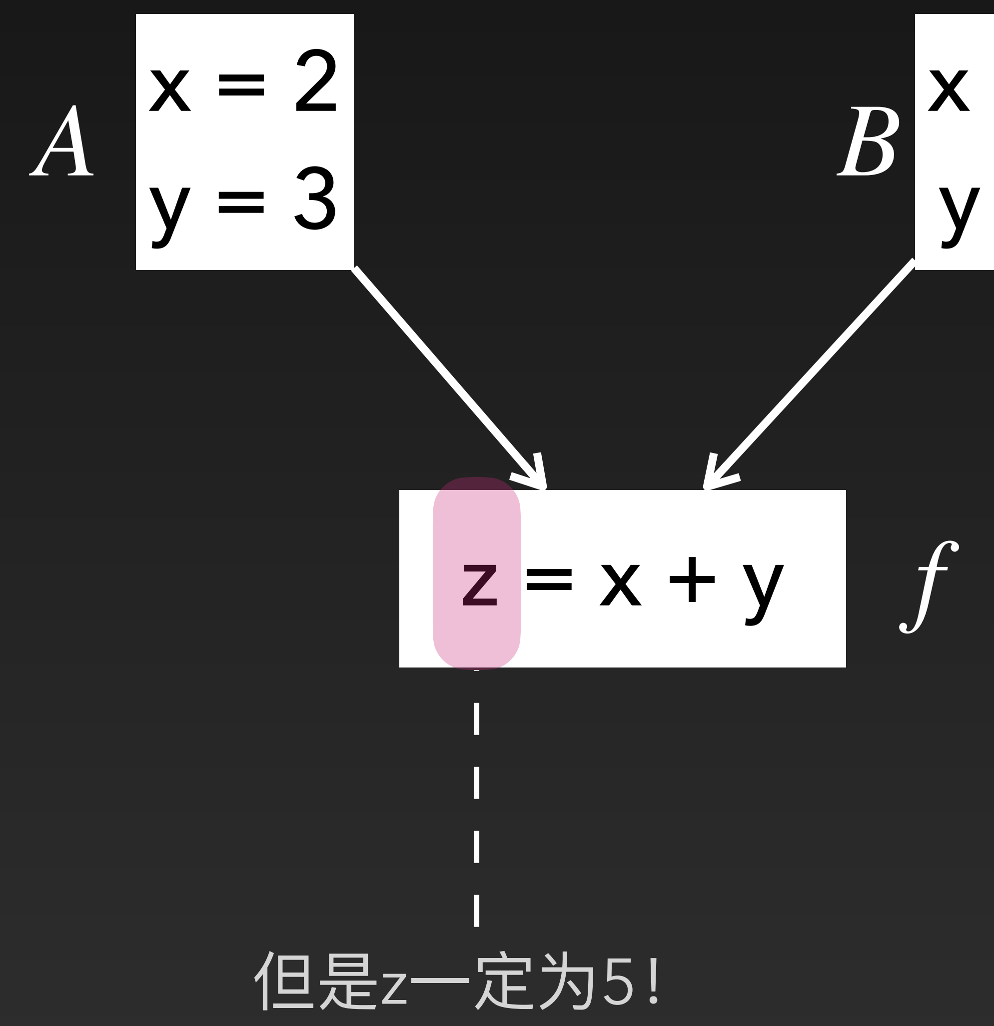


  - 虽然格是无穷，但符合ACC，即有限下降链，因此在其上的单调函数也是continuous的，可以用kleene算法计算不动点！
  - 此外，又含有 ⊤，⊥，加上ACC，因此也是一个完备格！

常量传播（Constant Propagation）

- 下面举一个简单的基本块内的语句：z = x + y，其转移函数

x	y	z
c1	NAC	NAC
	c2	c1 + c2
	undef	undef
undef	NAC	undef
	c2	undef
	undef	undef
NAC	undef	undef
		NAC



OUT[A] = {x = 2, y =3}, OUT[B] = {x = 3, y =2},

f (OUT[A]) = {z = 5, x = 2, y = 3}

f (OUT[B]) = {z = 5, x = 3, y = 2}

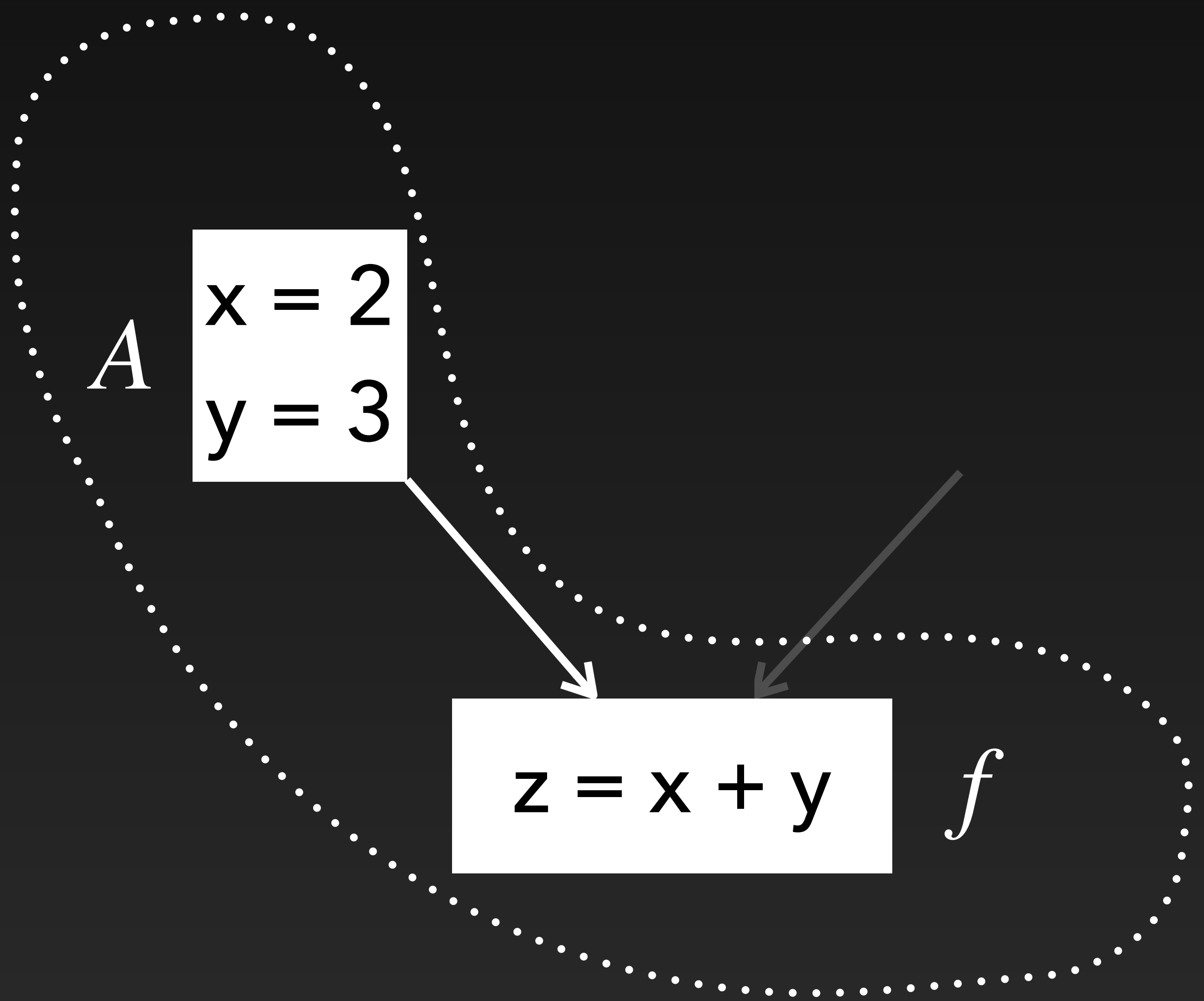
f (OUT[A] ∨ OUT[B]) = {z = NAC, x = NAC, y = NAC}

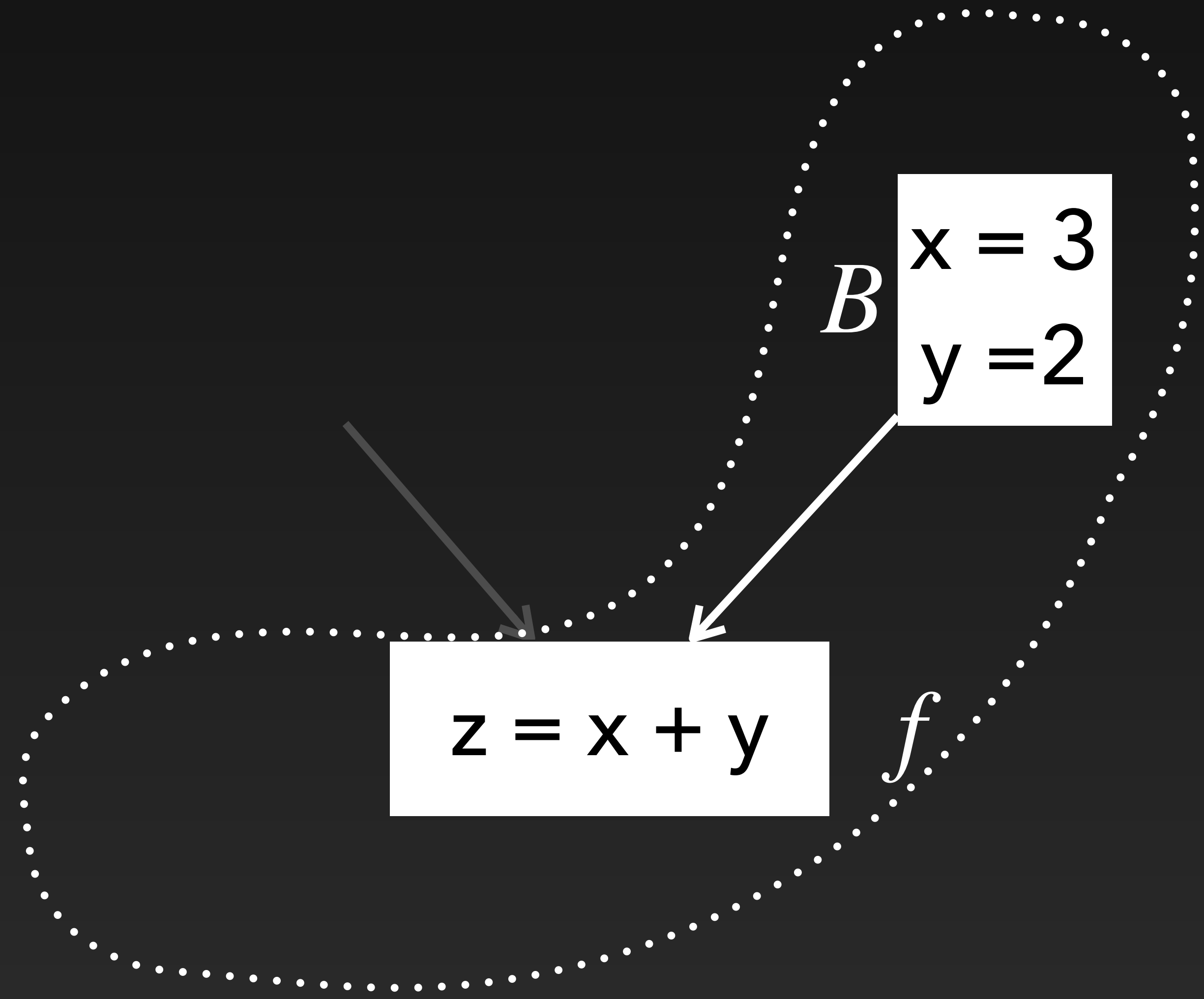
Path meet operator weakens the overall precise

常量传播（Constant Propagation）

- 下面举一个简单的基本块内的语句：z = x + y，其转移函数

x	y	z
c1	NAC	NAC
	c2	c1 + c2
	undef	undef
undef	NAC	undef
	c2	undef
	undef	undef
NAC	undef	undef
		NAC





OUT[A] = {x = 2, y =3}, OUT[B] = {x = 3, y =2},

f (OUT[A]) = {z = 5, x = 2, y = 3}

f (OUT[B]) = {z = 5, x = 3, y = 2}

f (OUT[A] ∨ OUT[B]) = {z = NAC, x = NAC, y = NAC}

Path meet operator weakens the overall precise

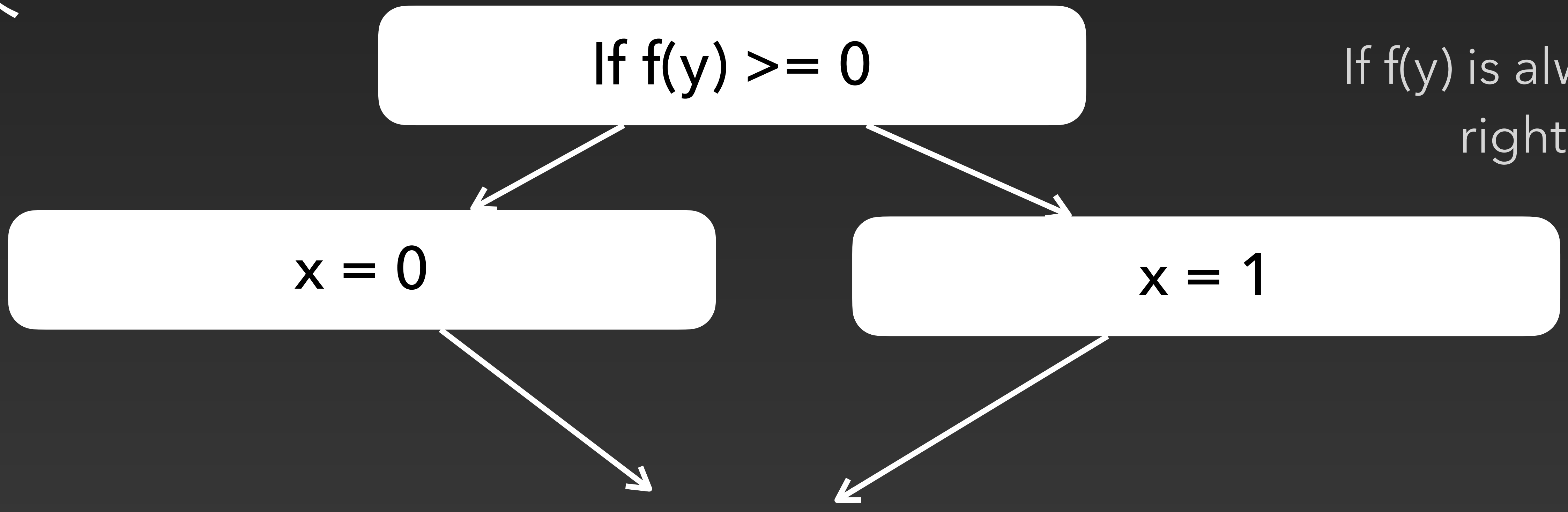
实际上一次的执行就是一条路径（path），这个信息只决定于这个路径上的前驱节点

如果我们想要的是“无论这个程序如何执行，这个点的信息都是某个值”，那么更加精确的表达是

f (OUT[A]) ∨ f (OUT[B]) = {z = 5, x = NAC, y = NAC}

完美的数据流分析

- 定义：给定$f_1, \ldots, f_m \in F$，其中f_i是第i个结点的转换函数
  - $f_p = f_{n_k} \cdot \ldots \cdot f_{n_1}$, 其中$p$是一条沿着结点$n_1, \ldots, n_k$的路径
  - 如果p是一条空的路径，那么f_p是一个全等函数
- 完美的数据流答案：
  - 对于每个结点n:
    - $\vee f_{p_i}(\top)$: p_i是一条从程序开始到达点n的路径，即完美的数据流需要考虑所有可达到结点n的路径的聚合结果



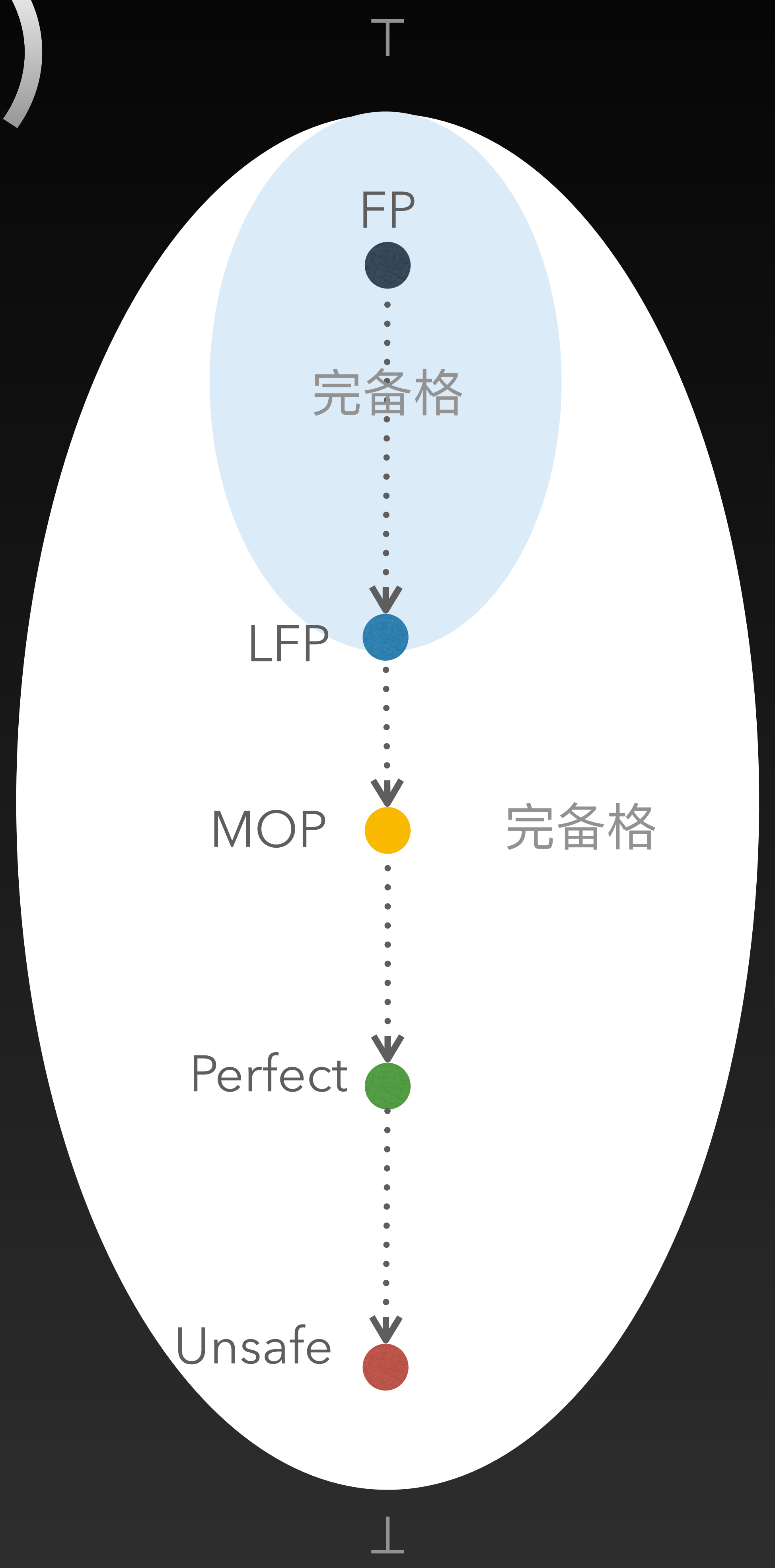
If f(y) is always non-negative then
right path never taken!

然而，事实上，决定所有可能的执行的路径是不可判定的！

路径汇聚(Meet-Over-Paths, MOP)

- Meet-Over-Paths(MOP)
  - 对于每个结点n: 对于所有可达n的路径p_i, $MOP(n) = \vee f_{p_i}(\text{init})$
  - 只要基本块存在边，就认为存在一条路径
    - 会过多考虑，因为有些路径是"死"路径
  - MOP = Perfect-Solution ∨ Solution-to-Unexecuted-Paths
    - Perfect-Solution ⪯ MOP

      这里⪯表达左边比右边精确
  - MOP是相比于Perfect-Solution一个更加小的但是更加安全的解
  - 但，MOP本身不可计算，因为可能有潜在无穷多的path（Loop），即你无法编程写出$\vee f_{p_i}(\text{init})$
- 我们有：$PERFECT \preceq MOP \preceq LFP \preceq FP$，策略： 尽可能和MOP接近



可分配框架（Distributive Framework）

- 一个框架$(F, V, \vee)$是分配的的，当且仅当
  - $f(x \vee y) = f(x) \vee f(y)$
  - 即输入的merge再apply函数 相等于 先各自apply函数，再merge
    - 比如定义可达分析，活跃变量分析
  - 我们在迭代算法中往往是做$f(x \vee y)$ （统一考虑所有可能输入），但实际上需要$f(x) \vee f(y)$ （即每种输入需要单独考虑）
  - 一般情况下，$f(x \vee y) \neq f(x) \vee f(y)$，这意味着你可能得到不那么精确的信息
    - 比如常量分析中 $x \leq x \vee y,\ y \leq x \vee y \implies f(x) \leq f(x \vee y), f(y) \leq f(x \vee y) \implies f(x) \vee f(y) \leq f(x \vee y)$

正确性和准确度

- 如果一个数据流分析框架是单调的，那么算法收敛时，有$MOP \leq LFP$
- 如果一个数据流分析框架是分配的，那么算法收敛时，$MOP = LFP$
  - meet-early (先聚合，再apply) = meet-late（先apply，再聚合，即MOP)
  - 定义可达和活跃变量的分析都符合要求
- 如果只有单调，但不分配
  - 那么$LFP \neq MOP$
  - 常量传播就是此类

抽象解释（abstract interpretation）

- 注意：LFP给出的是约束方程f (程序基本块、path meet/join的复合)的最小不动点了，而最小不动点已经代表了这个方程f 最为精确（有用）的解，但为什么还没有精确到perfect solution?
- 因为约束方程f的一种保守抽象，其不能完美的描述程序行为，而如果我们能够给出更加精确的约束方程f的描述，那么我们得到的LFP的分析就更加精确
- 怎么描述不同约束方程f的关系呢?
- 抽象解释就是系统研究这种关系的理论
  - 有机会之后来讲讲


Q&A